Naval Research Laboratory
Monterey, CA 93943-5502

NRL/MR/7543--97-7228

Aircraft Icing Algorithms Applied to U. S. Navy Numerical Model Data: A Verification Study

G. N. VOGEL

Meteorological Applications Development Branch
Marine Meteorology Division

February 1997

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. Agency Use Only *(Leave Blank)*.	2. Report Date.	3. Report Type and Dates Covered.
	February 1997	Final

4. Title and Subtitle.
Aircraft Icing Algorithms Applied to U. S. Navy Numerical Model Data: A Verification Study

5. Funding Numbers.
PE 0603207N
PN X1596-02

DN 153-164

6. Author(s).
G. N. Vogel

7. Performing Organization Name(s) and Address(es).
Naval Research Laboratory
Marine Meteorology Division
Monterey, CA 93943-5502

8. Performing Organization Reporting Number.
NRL/MR/7543--97-7228

8. Sponsoring/Monitoring Agency Name(s) and Address(es).
Space and Naval Warfare Systems Command (PMW-185)
Washington, D. C. 20363-5100

10. Sponsoring/Monitoring Agency Report Number.

11. Supplementary Notes.

12a. Distribution /Availability Statement.
Approved for public release; distribution unlimited.

12b. Distribution Code.

13. Abstract *(Maximum 200 words).*

The results from a verification with pilot reports of four aircraft icing algorithms applied to Navy global numerical model data are presented. Significant differences in forecast performance among algorithms are closely related to differences in temperature and moisture thresholds utilized to infer icing. Near 850 mb (~5000 ft), three of the icing routines correctly forecast over 70% of observed icing occurrences, and obtained Hanssen and Kuipers skill scores (difference between hits and false alarms) in excess of 0.5 (0, random performance; 1, perfect skill). Statistical tests indicated that differences in skill scores as a function of forecast lead time (0 to 24 hr) were generally not significant. The use of higher vertical resolution data was found very important for enhanced icing prediction performance. Overall results indicate that the ability of icing routines to differentiate icing type and intensity based on temperature, moisture and stability criteria is clearly limited. In terms of forecast skill and computational efficiency, algorithm comparisons indicate that the icing routine currently used as operational guidance at the National Centers for Environmental Prediction Aviation Weather Center would be a good selection for the Naval Research Laboratory's aviation support product suite.

14. Subject Terms.
Aircraft icing Icing algorithms NOGAPS

15. Number of Pages.
94

16. Price Code.

17. Security Classification of Report.	18. Security Classification of This Page.	19. Security Classification of Abstract.	20. Limitation of abstract.
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Special thanks are extented to Mr. Charles Sampson of NRL Monterey for his invaluable technical assistance and thorough review of this manuscript. The author wishes to thank personnel within the Research Applications Program at the National Center for Atmospheric Research, in particular Mr. Gregory Thompson, for graciously providing the PIREP data, and the RAP and AIRF icing algorithms, used in this study. This work is sponsored by the Oceanographer of the Navy through the Space and Naval Warfare Systems Command (PMW-185), under program element 0603207N.

AIRCRAFT ICING ALGORITHMS APPLIED TO U.S. NAVY NUMERICAL MODEL DATA: A VERIFICATION STUDY

1. INTRODUCTION

The Naval Research Laboratory (NRL) Marine Meteorology Division is presently developing an interactive information-processing system that will provide a suite of environmental products to operational users tasked with providing aviation weather forecast support. This automated display system is projected to significantly improve both flight safety and efficiency, by providing more accurate and timely warnings of weather impacted airspace to naval aviators. The aviation support product suite will provide graphical displays of both analyzed and forecast aviation impact variables derived from gridded numerical model data. For the initial product suite, an icing potential display is scheduled for implementation. This report evaluates several aircraft icing algorithms considered as viable candidates for selection as the icing potential product.

In this verification study, the results of a comparative evaluation of four different icing algorithms using gridded numerical model data from the Navy Operational Global Atmospheric Prediction System (NOGAPS) are presented. Specifically, the icing algorithms evaluated herein include those developed by the Research Applications Program of the National Center for Atmospheric Research (NCAR-RAP), the National Aviation Weather Advisory Unit (NAWAU), the Air Force and the Navy. Verification of the algorithms is accomplished by comparing model-derived analyses and forecasts of icing potential with pilot reports (PIREPs).

Forecasts of icing type and intensity provided by the Air Force and Navy algorithms are also statistically compared with reported conditions. In order to assess the impact of model accuracy and vertical data resolution on forecast performance, a second evaluation of algorithms is performed using coincidental model-derived profiles and radiosondes.

2. ICING ALGORITHMS

Operational numerical models such as NOGAPS do not explicitly forecast liquid water content or dropsize distribution which, along with temperature, are the key meteorological factors involved in aircraft ice accretion. Without appropriate cloud data, model-applied algorithms attempt to infer significant icing areas by locating those model temperature and humidity data within certain thresholds known to be favorable to icing potential. As formulated, icing algorithms do not consider aircraft and flight characteristics, important non-meteorological factors which significantly influence the rate of accumulation (i.e., intensity) of ice upon a particular airframe.

2.1 NAWAU

The NAWAU icing algorithm, developed by Ron Olson of the Aviation Weather Center's Aviation Observing Branch (formerly NAWAU), is used as operational guidance in the issuance of advisories to the aviation community (R.J. Olson, pers. comm., 1996). The algorithm, a refinement of one developed by Schultz and Politovich (1992), predicts two categories of icing (Table 1). Temperature (T) and relative humidity (RH) criteria for delineating icing-threat areas were determined through real-time

Table 1. Tabular description of the NAWAU icing algorithm, including temperature and relative humidity thresholds for both icing categories.

ICING CATEGORY	TEMPERATURE	RELATIVE HUMIDITY	HEIGHT
2 - HIGHER PROBABILITY*	$-20 \leq T \leq 0^{\circ}C$ $-14 \leq T \leq -1^{\circ}C$	RH $\geq$ 86% RH $\geq$ 75%	$\leq$ 900m above sfc > 900m above sfc
1 - LOWER PROBABILITY*	$-19 \leq T \leq 0^{\circ}C$	RH $\geq$ 60%	

* can be lowered by one category if downslope winds > -5 cm/s

Table 2. Tabular description of the RAP icing algorithm, including temperature and relative humidity thresholds for each icing category.

ICING CATEGORY	TEMPERATURE	RELATIVE HUMIDITY
FREEZING RAIN	$T \leq 0^{\circ}C$	RH $\geq$ 80% w/RH $\geq$ 80% above $T > 0^{\circ}C$
FREEZING DRIZZLE	$-12 \leq T \leq 0^{\circ}C$	RH $\geq$ 85% w/RH < 85% above $T < -12^{\circ}C$
UNSTABLE	$-20 \leq T \leq 0^{\circ}C$	RH $\geq$ 56% w/Max RH $\geq$ 65% below in conditionally unstable layer
STABLE	$-16 \leq T \leq 0^{\circ}C$	RH $\geq$ 63%

comparisons with PIREPs and satellite imagery by NAWAU forecasters. In an attempt to reduce the specified threat area (i.e., overforecasting), category 2 icing generally has more restrictive T and RH thresholds than category 1 icing. The only exception to this occurs at low levels (within 900 m of the surface) where the category 2 temperature threshold is expanded to include temperatures down to -20°C. Due to insufficient model (NOGAPS) data, the NAWAU algorithm feature which reduces areas of icing when orographic downslope flow exists, was not utilized in this evaluation study. The NAWAU icing algorithm has been previously evaluated by NCAR-RAP (Brown et al., 1994; Thompson et al., 1995).

2.2 RAP

The RAP icing algorithm, developed by NCAR's Research Applications Program, is a refinement of one developed by Forbes et al. (1993) and an extension of the Schultz-Politovich algorithm. As emphasized by Thompson et al. (1995), the RAP algorithm is continually evolving; the temperature and relative humidity thresholds for icing shown in Table 2 (and used in this study) represent those operative during the autumn of 1994. The algorithm consists of four categories of icing - freezing rain, freezing drizzle, unstable and stable. Although freezing precipitation is widely regarded as the most severe icing hazard, the four RAP algorithm categories are not designed to rank icing severity but rather provide different physical bases for icing diagnoses. For both freezing precipitation categories, there are T and RH criteria not only at the level in question, but also for a level above. The unstable (convective) icing category, in

addition to T and RH criteria at the level in question, requires the existence below of a conditionally unstable layer (i.e., lapse rate) with relative humidity $\geq$ 65%. The stable icing category is similar to the NAWAU category 1 icing. The version of the RAP icing algorithm described here has been evaluated in-house using data from several different models (Brown et al., 1994; Thompson et al., 1995). However, these evaluations have not specifically verified individual RAP icing category predictions (ex., freezing rain, drizzle) against actual reported occurrences.

2.3 AIRF

The Air Force (AIRF) icing algorithm was developed at the Air Force Global Weather Central as guidance for flight operations (Knapp, 1992). Originally written for application to radiosonde data, the AIRF algorithm uses temperature, dew point depression (T-Td) and stability (lapse rate) criteria to predict icing type and severity (Table 3). Specifically, given a below-freezing moist layer (T-Td $\leq 4^{\circ}C$) within a radiosonde or model profile, the AIRF algorithm assigns an icing type and intensity at each sounding (or model) level throughout that layer, based on the average temperature and dew point depression of the level and the next lower level, and the lapse rate between the two levels. For the bottom level of a moist layer, the algorithm always assigns the icing type and intensity corresponding to the next higher level. Three types of icing may be specified. For stable lapse rates ($\leq 2^{\circ}C$ / 1000 ft), rime (RME) is always specified; depending on the temperature, either clear (CLR) or

Table 3. Tabular description of the AIRF icing algorithm, including temperature, dew point depression and lapse rate criteria for specification of icing intensity and type. Forecast icing intensities and types are defined in the text.

TEMPERATURE (°C)	0≥T>-8				-8≥T>-16				-16≥T≥-22
DEW PT. DEPRESSION (°C)	≤1		1<T-Td≤2		≤1		1<T-Td≤3		≤4
LAPSE RATE (°C/1000ft)	≤2	>2	≤2	>2	≤2	>2	≤2	>2	N/A
FORECAST ICING	LGT RME	MDT CLR	TRC RME	LGT CLR	MDT RME	MDT MXD	LGT RME	LGT MXD	LGT RME

Table 4. Tabular description of the TESS icing algorithm, including temperature, dew point depression and lapse rate criteria for specification of icing probability, intensity and type. 'TBL' indicates intensity value determined using look-up table.

TEMPERATURE (°C)	DEW PT. DEPRESSION (°C)	LAPSE RATE (°C/100m)	PROBABILITY (%)	ICING TYPE	INTENSITY
5≥T>0	≤2	N/A	50	IND	UNK
0≥T≥-7	2≤T-Td<4	<.55	20	RME	TRC
		≥.55	20	MXD	TRC
	<2	<.55	100	RME	TBL
		≥.55	100	CLR	TBL
-7>T≥-15	3<T-Td<6	<.55	20	RME	TRC
		≥.55	20	MXD	TRC
	≤3	<.55	100	RME	TBL
		≥.55	100	MXD	TBL
-15>T≥-22	4≤T-Td<6	N/A	10	RME	TRC
	<4	<.55	100	RME	TBL
		≥.55	100	CLR	TBL
-22>T≥-30	<6	N/A	10	RME	TBL

mixed (MXD) icing is forecast for unstable conditions. Forecast icing intensities include trace (TRC), light (LGT) and moderate (MDT); heavy or severe icing is never specified. The AIRF icing algorithm has been previously evaluated (in terms of its ability to predict yes/no icing conditions) by Knapp (1992) and by NCAR-RAP (Brown et al., 1994; Thompson et al., 1995). A more comprehensive evaluation of the algorithm, which includes icing type and severity, has been performed by Cornell et al. (1995).

2.4 TESS

As an atmospheric application program within the Navy's Tactical Environmental Support System (TESS) , the aircraft icing probability function provides operational users at various land sites and aboard selected vessels within the fleet a remote site automatic icing analysis using radiosondes (Naval Oceanographic Office, 1988). This TESS icing algorithm is based on empirical forecast rules given in AWS/TR-80/001 (Air Weather Service, 1980) and updated cloud characterizations (Jeck, 1985). The algorithm provides a probability, type and intensity of aircraft icing for each level of a sounding (or model profile, if so chosen) based on various criteria detailed below. To date, the only documented evaluation of the TESS icing algorithm has been by Cornell et al. (1995).

Compared to the NAWAU, RAP and AIRF algorithms, the TESS icing algorithm is complex. Initially, the lifting condensation level (LCL) is computed, and this height is set as the base of the lowest cloud layer. When a superadiabatic layer exists at the surface, the LCL is not computed (actually, is not defined), and negative icing is specified at all levels of the sounding.

Based on temperature, dew point depression and stability (lapse rate) criteria, icing may be specified at a given sounding (or model profile) level within a cloud layer (Table 4). The icing probability can be 10, 20, 50 or 100%. (Note: at any given level, the probability of icing is zero if the level height is less than the cloud base height or Table 4 temperature and dew point depression icing thresholds are not met.) The icing type can be induction (IND), rime (RME) , mixed (MXD) or clear (CLR). Rime icing is forecast under stable conditions ($< 5.5^{o}C/km$), mixed or clear when the atmosphere is unstable or conditionally unstable ($\geq 5.5^{o}C/km$). For icing probabilities of 100% , the intensity of icing (TRC, LGT, MDT or SVR (severe)) is determined from a look-up table, and is based on the cloud temperature, the icing type and the distance between the cloud base height and the level being analyzed. Icing intensity is specified as unknown (UNK) for induction icing, and trace at probabilities of 10 to 20% when the temperature is between 0 and $-22^{o}C$. The TESS icing algorithm permits multiple cloud (viz., icing) layers to be specified for a given sounding. A new cloud layer is indicated when icing is forecast to occur, the icing probability at the previous lower level was zero, and the air temperature is $< 5^{o}C$.

The threshold values and resultant icing conditions shown in Table 4 (and used in this study) represent those from an actual operational version (i.e., computer code) of the TESS algorithm. In some cases, Table 4 temperature and dew point depression category endpoints do not agree with those given in the TESS 2.0 program performance specification for aircraft icing

(Naval Oceanographic Office, 1988). Additionally, the Table 4 icing type for $-15 \leq T < -7^{\circ}C$ and 100% probability is mixed; according to Naval Oceanographic Office (pg. 491), the type should be clear. For this verification study, the TESS algorithm operator-selected option of specifying whether or not a cloud layer intersects a frontal inversion was set to "no", thereby eliminating the need of the algorithm's second icing intensity look-up table used when a frontal inversion is present.

3. DATA

3.1 Model - NOGAPS

The Navy Operational Global Atmospheric Prediction System provided the model data required for the computation of the icing algorithms evaluated in this study. The NOGAPS is run twice daily (at 00Z and 12Z) at the Fleet Numerical Meteorology and Oceanography Center (FNMOC). Data for this study is from the NOGAPS Version 3.4 forecast model which consists of a multivariate optimum interpolation analysis, a nonlinear normal mode initialization scheme, and a 159-wave triangular (~ 3/4 deg. horizontal resolution), 18 -level spectral forecast model (Hogan et al., 1991; Goerss and Phoebus, 1993).

During the two and a half month period from mid-March through May 1995, NOGAPS temperature, geopotential height and vapor pressure fields, interpolated onto a $2.5^{\circ} \times 2.5^{\circ}$ spherical grid, were stored using the Naval Environmental Operational Nowcasting System (NEONS) database. As available, meteorological data fields were archived at three forecast lengths (the analysis, and 12 and 24 hours) and at seven constant pressure surfaces (1000, 925, 850, 700, 500, 400 and 300 mb). Analysis

data represents a blend of a 6 hr background forecast with quality-controlled current observations, performed prior to model initialization. Simultaneous with each NEONS data extraction, gridded relative humidity fields (at the same resolution) were computed and archived. NOGAPS data were collected for a regional grid field encompassing the continental U.S. and adjacent areas (viz., $5^{o}N$-$60^{o}N$, $50^{o}W$-$130^{o}W$).

3.2 PIREPs

3.2.1 Description

For the period March through May 1995, a database comprising over 119,000 aircraft pilot reports was obtained from the National Center for Atmospheric Research. These PIREPs contain encoded numeric data which provide information on weather, cloud layers, turbulence and icing. For each PIREP, up to two icing layers or levels may be specified, each as a coded group containing icing base and top heights, intensity and type. Additional icing information may be available within a PIREP as alphanumeric remarks.

The base and top heights of a reported icing layer are usually specified to the nearest thousands of feet, although some heights are reported to the nearest hundreds of feet. For those PIREPs indicating clear skies or clear above (current flight level), the top height of any reported negative icing layer is always specified (i.e., coded) as 60,000 ft. Icing type is encoded as either negative, rime, clear or mixed. Additional code elements were created for icing due to freezing rain (ZR) and freezing drizzle (ZL), with such occurrences only ascertained

through pilot remarks. Icing intensity is defined numerically as follows: 0) negative, 1) trace, 2) trace-light, 3) light, 4) light-moderate, 5) moderate, 6) moderate-heavy, 7) moderate-severe and 8) severe. Within a given PIREP, a numeric value of '-9' is assigned any icing code element (intensity or type) for which no information is given.

3.2.2 Selection Procedure

Various criteria were applied to the original PIREP database to select those pilot reports finally used as verification data for model-derived icing forecasts. PIREPs needed to be located within the study's model grid domain and were required to have occurred within one hour of 00Z or 12Z. All selected PIREPs had to provide at minimum an icing type or intensity for a specified level or layer. Given model data at the appropriate date and time, a PIREP was matched to the grid point nearest to the PIREP, provided the report was no more than 80 km from that model grid point, and at an elevation below a specified top height (~ 27,000 ft). For each PIREP, the top height (for icing) was specified as the average model analysis (or, if unavailable, the 12 or 24 hr forecast) height of the 300 and 400 mb pressure surfaces at the nearest gridpoint. For some PIREPs, pilot remarks provided new or additional information which was numerically encoded provided such information met location and time criteria.

Having passed the above selection criteria, a PIREP's icing information was assigned (based on reported elevations) to one or more of the following five icing layers: LOW, LOW-MID, MIDDLE, MID-HIGH and HIGH. The midpoints (in elevation) of the LOW-MID, MIDDLE, MID-HIGH and HIGH layers correspond to the 850, 700, 500

and 400 mb model constant pressure surfaces, respectively; the lowest icing layer (LOW) encompasses both the 1000 and 925 mb model constant pressure surfaces. For each PIREP, the base (top) height for the highest four layers is specified as the average model height of the layers's constant pressure surface and the next lower (higher) pressure surface at the chosen gridpoint. For example, the base (top) height of the MIDDLE icing layer at a given gridpoint is the average of the 850 and 700 mb (700 and 500 mb) model heights. The top of the LOW icing layer is specified as the average height of the 925 and 850 mb pressure surfaces; the surface represents the base of this layer. PIREPs within 5 minutes and 40 km of each other were retained as separate reports if their icing information was from different layers or, combined into one report, if they provided information for the same layer. In general, for any PIREP (individual or combined) reporting two different icing intensities within a single layer, the larger intensity value is selected for comparison with model-derived icing forecasts. The icing type for a layer is that which corresponds to the chosen (i.e., highest) intensity value. In those cases where a PIREP has two equal maximum icing intensities with differing icing types within a single layer (ex. MIDDLE), the type 'mixed' is specified for model comparisons.

3.2.3 Distribution

Applying the selection criteria outlined above, a dataset comprised of 1757 PIREPs over the period mid-March through May 1995 was created for comparison with icing algorithm forecasts

derived from NOGAPS 2.5° data. Included within this PIREP dataset are 11 reports of icing type only (i.e., no intensity). Temporally, about 56% of the pilot reports correspond to 12Z, the remainder to 00Z. The spatial distribution of the PIREP dataset, and the number of negative icing reports, per 5° x 5° latitude-longitude bins, is shown in Figure 1. The largest concentration of PIREPs in the dataset occurs over the area encompassing the southern Great Lakes and northern Ohio Valley. Secondary maxima in number of reports occur over eastern Colorado and western Washington. While the total number of negative icing reports is largest in the two 5°x 5° bins with the largest number of total PIREPs, the highest frequency of negative icing reports occurs over the south-central states, with 108 of 139 (~78%) PIREPs between 25-35°N and 90-105°W classified as negative icing reports. Overall, about 41% of the PIREPs within the dataset are negative icing reports.

Figure 2 presents the number of positive icing reports, per 5° x 5° latitude-longitude bins, of intensity less than moderate or greater (TRC, TRC-LGT or LGT) and of intensity moderate or greater (LGT-MDT and above). For this compilation, the icing value assigned any individual PIREP corresponds to the maximum reported intensity. The largest concentration of positive icing reports occurs over the southern Great Lakes - northern Ohio Valley area. Over the western U.S., positive icing reports are largely concentrated over the central Rockies and Pacific Northwest. Reports of icing conditions are not common south of 35°N. Of any 5°x 5° box with greater than 7 positive icing reports, only one (that containing Pennsylvania and western New York)

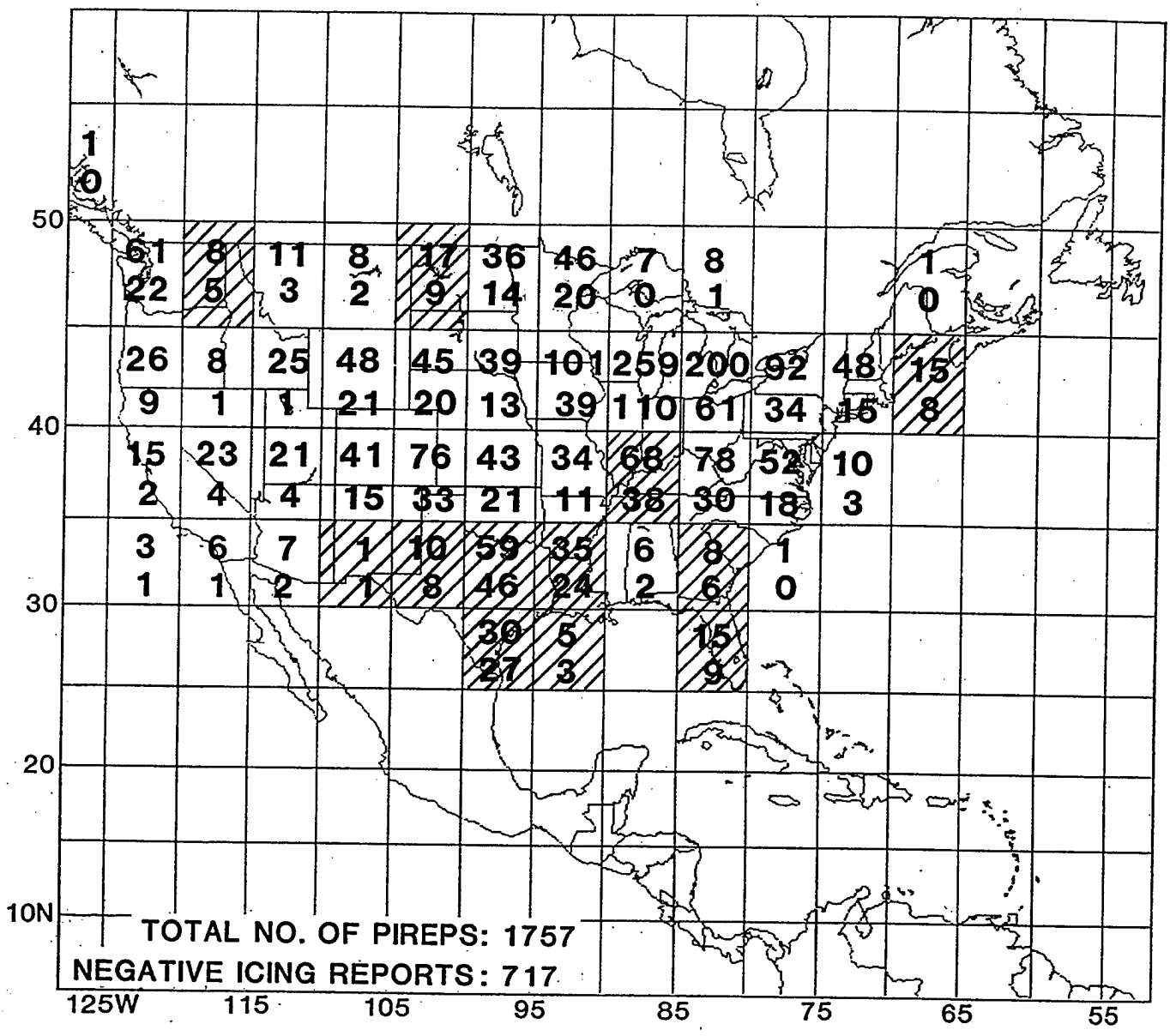

Figure 1. Spatial distribution of the 2.5° PIREP dataset, corresponding to the period March-May 1995. For 5° x 5° latitude-longitude boxes, the total number of PIREPs and the number of negative icing reports are given. Striped boxes indicate ≥50% negative icing reports.

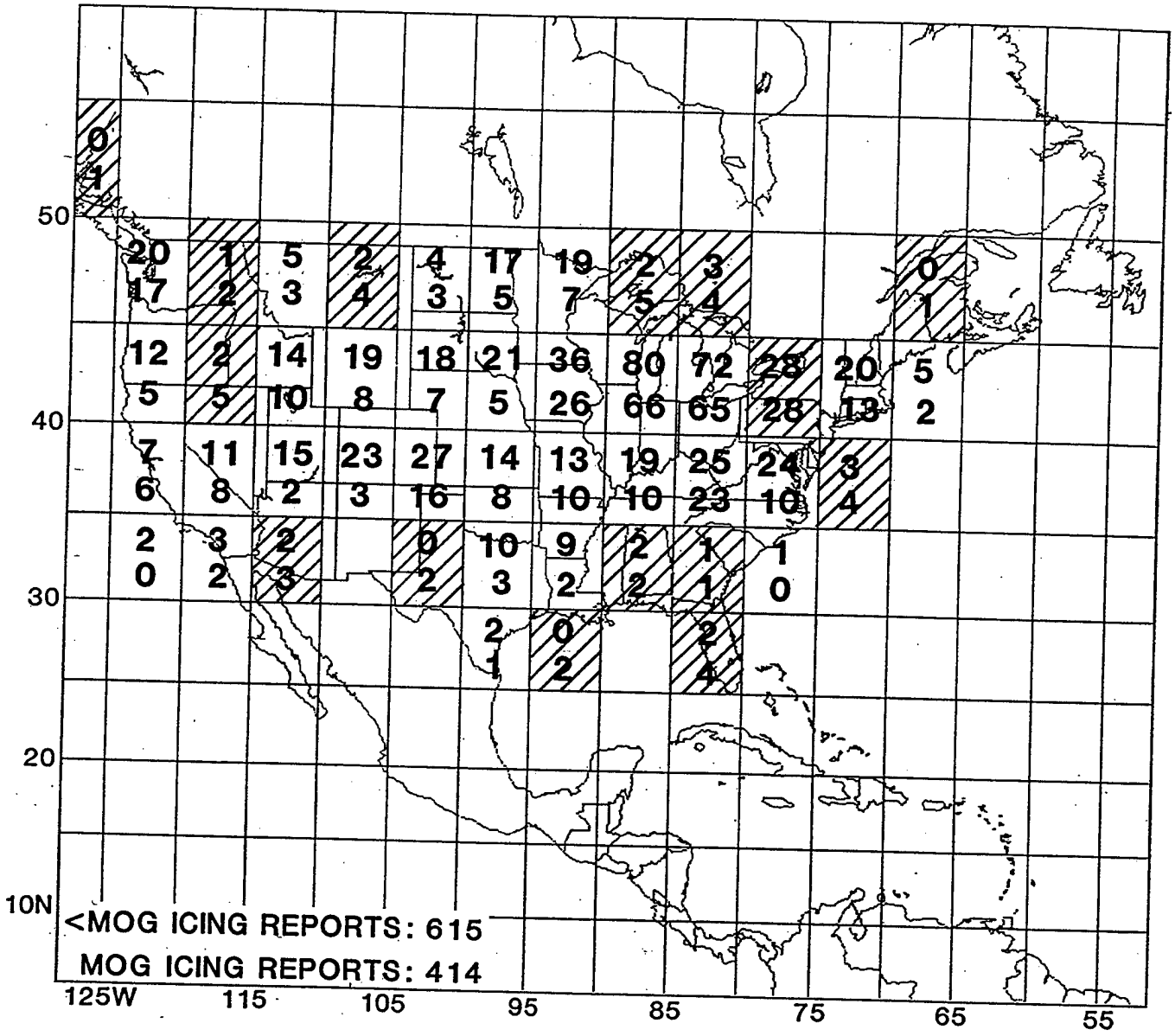

Figure 2. Spatial distribution of positive icing reports for the 2.5° PIREP dataset. For 5° x 5° latitude-longitude boxes, the top number corresponds to <MOG intensity reports and the bottom number to MOG intensity reports. Striped boxes indicate the number of MOG reports ≥ number of <MOG reports.

has at least as many MOG (moderate or greater) icing reports as less than MOG reports (28 each category). This box, plus the three to its west and the one to its southwest, contain half of all MOG intensity reports within the dataset. Overall, about 40% of all positive icing reports are of MOG intensity.

Characteristics of the 2.5° PIREP dataset based on vertical icing layers are given in Table 5. Due to the availability of icing (and no icing) information in multiple layers with many PIREPs, the total number of (layer) reports (3422) is roughly twice as large as the number of PIREPs included with this dataset. Of the five specified icing layers, the MIDDLE icing layer contains the most reports (about a third) while the LOW layer contains the least (< 7%). Overall, about 5 of every 8 layer reports are classified as negative icing reports. The total number of such reports (2144) is three times the number of negative icing PIREPs, indicating that many of these PIREPs provided multiple layer reports as a result of low-level flight under clear sky (no cloud) conditions. The percentage of negative icing reports as a function of all reports is about 48% at the MIDDLE layer; this percentage increases both upward and downward from this layer, reaching 63% of all reports at the LOW layer and 94% at the HIGH layer. When icing is specified, about three of every four reports indicate type 'rime'. Mixed (clear) icing is specified about once in 5 (15) reports. Almost three of every four reports of clear icing correspond to the LOW-MID and MIDDLE layers. All reports of freezing precipitation (ZR or ZL) occur at or below the MIDDLE icing layer. Slightly less than

Table 5. Characteristics of the 2.5° PIREP dataset, including the number of reports according to icing type and intensity, and the average base and top heights (in ft), for selected icing layers.

	LOW 925 MB	LOW-MID 850 MB	LAYER MIDDLE 700 MB	MID-HIGH 500 MB	HIGH 400 MB	TOTAL
AVG. HGT.						
BASE	----	3592	7292	14167	21103	----
TOP	3611	7269	14081	21110	27043	----
NO.RPTS.	228	703	1154	796	541	3422
TYPE						
NO (NEG.)	144	403	553	536	508	2144
RIME	48	186	425	184	27	870
CLEAR	9	20	37	11	0	77
MIXED	21	62	98	42	3	226
FRZ. DRZL.	0	1	0	0	0	1
FRZ. RAIN	1	3	2	0	0	6
NOT GIVEN	5	28	39	23	3	98
INTENSITY						
NEGATIVE	144	403	553	536	508	2144
TRACE	9	35	53	21	4	122
TRC-LGT	1	2	0	0	0	3
LIGHT	36	137	290	132	21	616
LGT-MDT	12	31	68	31	1	143
MODERATE	22	79	173	74	7	355
MDT-HEAVY	1	5	1	1	0	8
MDT-SEVERE	1	2	2	0	0	5
SEVERE	0	3	9	0	0	12
NOT GIVEN	2	6	5	1	0	14

half of all positive icing PIREPs specify the intensity as 'light' (categories TRC-LGT and LGT). When combined, the intensity categories LGT-MDT and MDT account for about 2 out of every 5 intensity reports. Trace icing is specified in slightly less than 10% of the reports. Heavy/severe icing (categories MDT-HEAVY, MDT-SEVERE and SEVERE) is seldom reported (only 2% of all layer reports); the LOW-MID and MIDDLE layers account for 88% of all reports of such icing.

Within the 2.5° PIREP dataset, a total of 1166 layer reports provided both icing type and intensity information. Table 6 presents the number of occurrences (and percent frequencies) of selected (intensity,type) pairs. Here, the top number of any (intensity,type) box is the number of occurrences, the middle value is the percent frequency of the specified type category for the given intensity and the lower value is the percent frequency of the specified intensity category for that given icing type. Results indicate that the percent frequency of occurrence of rime icing decreases somewhat with increasing intensity, while the reverse occurs for clear icing. The percent occurrence of mixed icing does not vary greatly over icing intensity categories (17% to ~22%). The percent frequency of icing intensity is rather similar for the three main icing types (RME, MXD and CLR). For example, depending on which icing type is chosen, between 46% and 51% of all its type observations occur at the light (TRC-LGT, LGT) intensity category. With only 4 available data, no conclusions can be drawn for (intensity,type) relationships for icing due to freezing precipitation (ZL,ZR).

A comparison of this study's dataset icing type and inten-

Table 6. Characteristics of the 2.5° PIREP dataset, according to selected intensity/type pairings. For any intensity/type box, the top number is the number of occurrences, the middle value is the percent frequency of the specified type category given the intensity, and the lower value is the percent frequency of the specified intensity category given the icing type.

INTENSITY \ TYPE	RME	MXD	CLR	ZL,ZR	TOTAL
TRC	74 78.7% 8.6%	16 17.0% 7.2%	4 4.3% 5.3%	0 0.0% 0.0%	94
TRC-LGT LGT	442 75.8% 51.2%	102 17.5% 45.7%	37 6.3% 48.7%	2 0.3% 50.0%	583
LGT-MDT MDT	333 71.5% 38.6%	101 21.7% 45.3%	31 6.7% 40.8%	1 0.2% 25.0%	466
MDT-HVY MDT-SVR SVR	14 60.9% 1.6%	4 17.4% 1.8%	4 17.4% 5.3%	1 4.3% 25.0%	23
TOTAL	863	223	76	4	1166

sity frequencies with various other observational datasets available in the literature is presented in Table 7. The icing type 'rime' is found to be dominant in all datasets, with frequencies ranging from 72% to 84% (Table 7a). Although the frequency of mixed icing reported by Cohen (1983) is about half of what is reported in the other studies, mixed icing is reported to be more common than clear icing in all four comparison datasets. For 6 of the 7 tabulated datasets, the icing intensity category 'light' is the most dominant frequency. It should be pointed out that the AWS (1980) data, adapted from all-weather flight test data compiled by Thompson (1955, Figure 16 of report), represents icing severity categories one less than originally given (ex., light instead of moderate). Eliminating the Perkins et al. (1957) dataset (where the categories trace and light are combined), reported frequencies of light icing range from about 43% to 67% among the given studies. The dataset used by Cornell et al. (1995) is unique in that almost all pilot reports are of either trace (the dominant frequency at 54%) or light intensity, and few of moderate or greater severity. Frequencies of moderate icing are observed to be quite large for this study's dataset and that utilized by Brown et al. (1994); in both of these studies, similar PIREP data processing procedures were followed. Excluding the Thompson (1955) dataset, observations of heavy/severe icing are rare, occurring at frequencies between zero and about 3%. Note that the frequency of heavy icing would be only 2.2 % for the Thompson dataset if, as per Air Weather Service (1980), a distinction is made between heavy and severe

Table 7. Percent frequencies of occurrence of icing types (a) and intensities (b) for this study's PIREP dataset compared to frequencies from various other observational datasets found in the literature.

(a)

NO. DATA	TYPE RME	MXD	CLR	OTHER	SOURCE/REFERENCE
1180	73.7%	19.2%	6.5%	0.6% ZL,ZR	Tbl. 5/This study
9693	78.1%	15.8%	6.1%	——[1]	Tbl. 4/Cornell et.al.,1995
114	84.2%	8.8%	7.0%	——[1]	Tbl. 8/Cohen,1983
4600	72.0%	17.0%	10.0%	1.0% FROST	Perkins et.al.,1957

Notes: 1 No Other Type

(b)

NO. DATA	INTENSITY TRC	LGT	MDT	HVY/SVR	SOURCE/REFERENCE
1264	9.7%	49.0%	39.4%	2.0%	Tbl. 5/ This study
9693	53.7%	42.8%	3.4%	——[1]	Tbl. 4/ Cornell et.al.,1995
40955	4.4%	56.3%	36.2%	3.2%	Fig. 4b/ Brown et.al.,1994
114	14.0%	67.5%	18.4%	0.0%	App. A/ Cohen,1983
337[2]	38.4%	44.7%	16.4%	0.6%	Tbl. 2/ Air Wea. Serv.,1980
——[3]	——[4]	87.0%	12.0%	1.0%	Perkins et.al.,1957
368	——[1]	48.9%	35.1%	16.0%	Fig. 9/ Thompson,1955

Notes: 1 Not specified
2 Based on Dewpoint Spread < 3°C
3 Unknown
4 Category combined with LGT

categories , with each category lowered one intensity level.

3.3 Radiosonde

In addition to NOGAPS 2.5° model data and pilot reports, radiosonde observations (RAOBs) were archived during the study period for the purpose of assessing the effect of model accuracy and resolution on icing forecasts. FNMOC quality-controlled RAOBs were obtained from the NEONS database. Typically, RAOBs provided geopotential height, temperature, dew point temperature and wind at mandatory and significant levels. The 925 mb constant pressure surface, a (NOGAPS forecast) level for this study, is not a mandatory radiosonde level; under fortuitous circumstances, information for this level may be provided in the significant level data. RAOBs do not directly provide the relative humidity; this parameter, required by the NAWAU and RAP icing algorithms, was derived subsequently during data processing. For the period March through May 1995, over 8,000 RAOBs from 85 sites (both civilian and military) within the continental U.S. were archived.

As was done with the PIREPs, selection criteria were applied to RAOBs to determine those eventually used for data comparisons. For each of the 1757 previously selected PIREPs, a search was made to find a RAOB at the appropriate date and time (either 00Z or 12Z) and within 80 km of the PIREP location. If successful, the distance from the grid point location previously chosen for the PIREP to the radiosonde site was calculated and, if within 80 km, the radiosonde was selected for study comparisons. The rather severe requirement that the PIREP, model gridpoint and radiosonde locations all be within 80 km dramatically

reduced the final database to only 280 RAOBs at 25 sites (see Figure 3). It should be pointed out that each of these RAOBs is not unique; in some cases, the same RAOB was assigned to multiple PIREPs. Geographically, the locations of radiosonde sites are fairly well distributed throughout the continental U.S., although no sites are available for the Northeast and within a rather wide swath from the northern Rockies and Great Plains southeastward to the mid-South (viz., Georgia). By far, the largest number of radiosondes used for model/PIREP data comparisons is at Denver, CO (67, or ~24%); this site, plus Pittsburgh, PA, account for about 3 of every 8 radiosondes utilized. In addition to radiosonde locations, Figure 3 provides the total number of PIREPs assigned each RAOB site according to three categories (negative icing, < MOG intensity and MOG intensity); one PIREP (over Oregon) did not report an icing intensity. Overall, the percentage of < MOG intensity reports used for radiosonde comparisons (~35%) is about the same as the percentage of such reports within the full PIREP dataset, while the percentages for negative icing and MOG intensity reports are slightly higher and lower, respectively. Although the number of MOG intensity reports is largest at the Denver, CO radiosonde site (14), the largest percentage of such reports at any one site (50%, or 7 of 14) occurs at Reno, NV.

For direct comparisons with PIREPs and model data, individual radiosondes are subdivided into the same five vertical layers - LOW, LOW-MID, MIDDLE, MID-HIGH and HIGH - as used for pilot reports. The top and bottom heights for these layers are found in

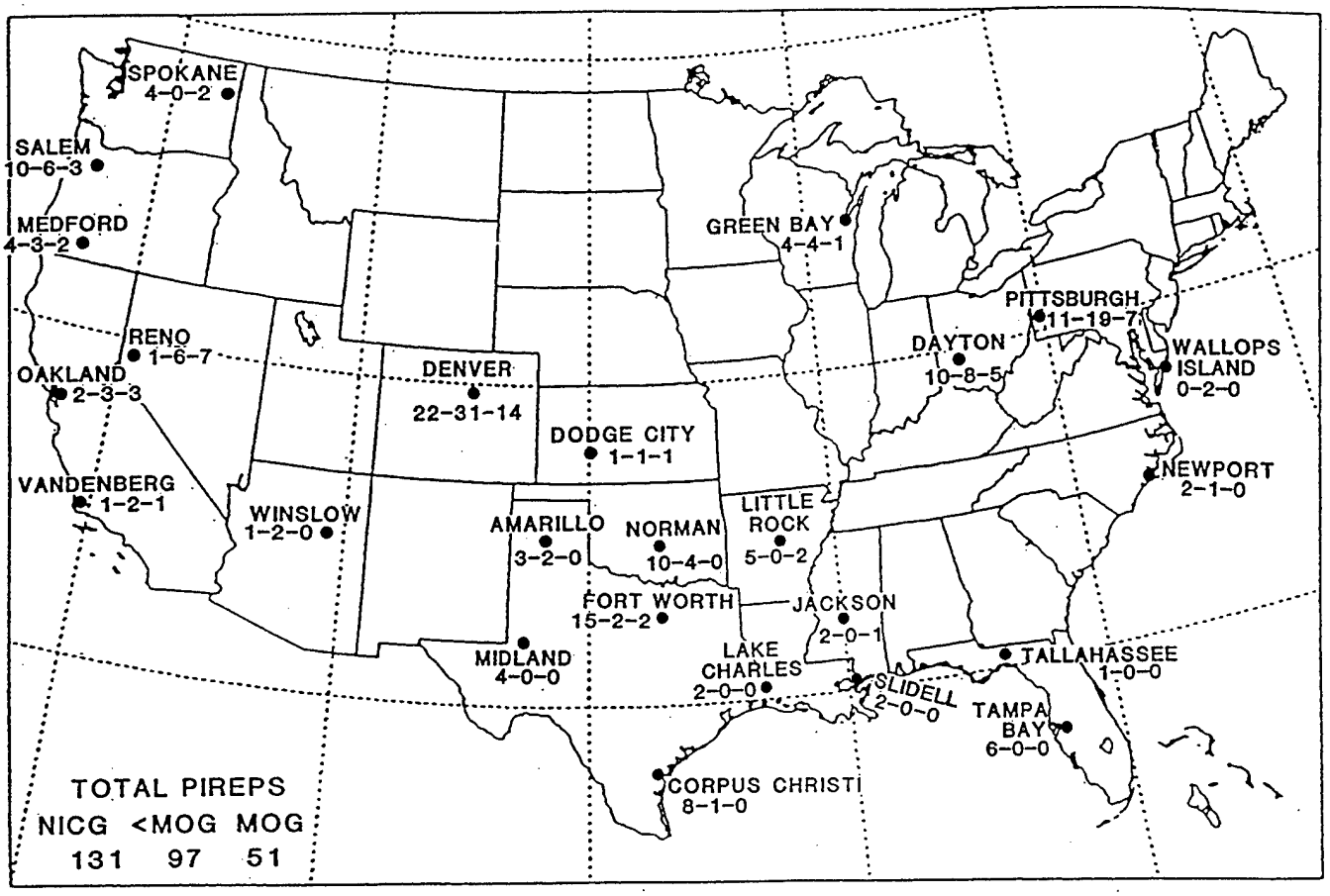

Figure 3. Radiosonde locations used for study comparisons. The number of PIREPs assigned each radiosonde site are indicated according to three categories: negative icing, <MOG intensity icing, and MOG intensity icing (left to right, respectively).

an analogous manner as those for PIREPs, except that radiosonde (mandatory) heights are used in lieu of model constant pressure surface heights to set layer limits. When needed, a radiosonde height value at 925 mb (and at any missing mandatory level, surface to 300 mb) is computed by logarithmically interpolating with respect to pressure using height values of bracketing isobaric surfaces. Application of any of the four icing algorithms used in this study to a particular radiosonde will provide a yes/no icing determination for each vertical layer with at least one radiosonde level. For any layer with multiple radiosonde levels, a positive icing condition is specified if icing is diagnosed at one or more levels. For radiosonde-derived AIRF or TESS icing analyses, the specification of an icing intensity and type for any given radiosonde layer containing data at two or more levels is analogous to the procedure used for PIREP layers containing two different intensities or types. In some instances, the TESS icing type for a layer reporting two equal maximum intensities is determined by comparing the forecast icing probabilities of the two intensities, then selecting the icing type corresponding to the larger probability.

3.4 Model Versus Radiosonde

Since model-applied algorithms infer icing by locating temperature and humidity data within certain critical thresholds, forecasts of icing may be significantly affected by biases in model state parameters. If one assumes radiosonde measurements are accurate to a certain degree of precision, then average model minus radiosonde differences (i.e., model biases) may be considered to be significant if the magnitude of such differences is

well in excess of known radiosonde instrument precision. For the 25 radiosonde sites shown in Figure 3, the VIZ 'B-sonde' was the dominant instrument type. According to Ahnert (1991), this radiosonde has the following measurement precision : 2 mb, or about 15-20 m (pressure, height), 0.3°C (temperature) and 1.6 % (relative humidity). Within the lower troposphere (below 400 mb), Ahnert finds that radiation errors are insignificant for the VIZ 'B-sonde.'

Table 8 provides average radiosonde values and model biases for geopotential height, temperature, dew point temperature and relative humidity, at seven constant pressure levels and three forecast lengths (Tau = 0, 12 and 24 hr). These statistics are based on 280 radiosondes, some of which are not unique (due to multiple PIREPs for the same sonde). The number of data at lowest level (1000 mb) is substantially reduced due to a predominance of radiosonde launches at higher elevations. Data availability at 925 mb is lacking since this pressure level is not a standard radiosonde reporting level. Below 850 mb, statistical sample size was further reduced due to significant differences in radiosonde and model elevations at several sites. Sample size was largest (smallest) at Tau = 12 hr (0 hr) because more (less) model data were available. Geopotential height biases are all well below radiosonde instrument precision (15-20 m). Eliminating the data-deficit 1000 mb level, temperature biases roughly twice instrument precision (i.e., between 0.5 and 0.9°C) only occur between 925 and 700 mb at Tau = 12 and 24 hr. These 'too warm' temperature biases are related, by hydrostatic

Table 8. Average radiosonde values and NOGAPS 3.4 model biases for geopotential height, temperature, dew point temperature and relative humidity, at seven pressure levels and three forecast lengths (Tau = 0, 12 and 24 hr).

		TAU = 0		TAU = 12		TAU = 24	
PRES. (MB)	NO. DATA	AVG. RAOB	BIAS	AVG. RAOB	BIAS	AVG. RAOB	BIAS
				GEOPOTENTIAL HEIGHT (M)			
1000	18,22,19	115	-2.4	109	-2.0	115	-2.0
925	42,45,42	756	0.8	755	2.4	758	0.1
850	178,187,182	1460	-0.6	1458	2.4	1460	0.8
700	241,257,253	3022	-0.6	3023	1.7	3025	1.2
500	240,255,251	5614	-0.5	5613	5.7	5616	5.4
400	240,256,252	7239	-0.5	7238	6.9	7241	7.1
300	237,253,249	9222	-2.1	9219	4.7	9223	5.3
				TEMPERATURE (°C)			
1000	18,22,19	19.3	1.9	18.7	-0.7	19.5	-0.5
925	42,45,42	11.7	0.2	11.8	0.7	11.8	0.6
850	178,187,182	5.3	0.1	5.5	0.7	5.5	0.9
700	241,257,253	-2.3	-0.1	-2.4	0.5	-2.4	0.5
500	240,255,251	-18.6	-0.1	-18.7	0.2	-18.6	0.2
400	240,256,252	-30.4	-0.0	-30.5	0.0	-30.5	0.0
300	237,253,249	-44.7	-0.2	-45.0	-0.2	-44.9	-0.1
				DEW POINT TEMPERATURE (°C)			
1000	18,22,19	16.7	-1.0	16.1	-0.3	17.0	-0.3
925	42,45,42	8.0	0.4	8.3	0.7	8.0	1.4
850	178,187,182	1.0	-0.9	1.1	-0.4	1.2	-0.4
700	241,257,253	-10.9	1.5	-10.7	2.1	-10.7	2.3
500	240,255,251	-30.3	1.5	-30.5	2.5	-30.5	2.8
400	240,256,252	-42.4	0.9	-42.6	2.2	-42.5	2.3
300	237,253,249	-56.2	-2.4	-56.5	-1.0	-56.4	-1.1
				RELATIVE HUMIDITY (%)			
1000	18,22,19	85.4	-13.7	85.5	1.8	86.5	1.0
925	42,45,42	83.4	-1.0	84.4	-2.3	82.9	2.5
850	178,187,182	79.6	-4.2	78.8	-5.7	79.6	-5.7
700	241,257,253	61.8	5.3	63.2	5.6	63.1	6.6
500	240,255,251	43.9	4.6	43.7	8.0	43.4	9.4
400	240,256,252	36.2	3.4	36.1	9.6	35.8	10.7
300	237,253,249	29.4	-6.7	29.3	-0.9	29.1	-1.6

consistency, to slightly positive geopotential height biases (5 to 7 m) found at higher levels (viz., 500 to 300 mb). By far, the most significant model biases, and those most likely to impact icing algorithm forecast performance, occur in moisture. Between 700 and 400 mb, the model is considerably more moist than what is observed, with biases increasing steadily with forecast lead time. This 700-400 mb layer represents a well-defined, distinct moist layer; Td and RH biases immediately below and above (at 850 and 300 mb, respectively) are all negative, indicative of a model atmosphere that is too dry. A very large negative RH bias occurs at Tau = 0 hr and 1000 mb; the significance of this statistical value is unclear since it is based only on 18 data and does not have a corresponding large dew point temperature bias.

4. VERIFICATION TECHNIQUES

Two statistical indices and a skill score discriminant, derived from two by two contingency tables, are used to evaluate the ability of a chosen icing algorithm to predict discrete yes/no icing conditions. Given a particular event (viz., icing), the prefigurance PF (also known as the hit rate or Power of Detection) is the capability of correctly forecasting that event, and is defined as the number of correct (model- or RAOB-derived) forecasts divided by the number of reported occurrences. The false alarm rate FAR is a measure of the tendency to overforecast, and is defined as the number of incorrect forecasts divided by the number of reported "no icing" occurrences. The FAR index can be thought of as the probability that a 'no icing' event will be incorrectly forecast. Due to the nonsymetric,

biased nature of the PIREP database, the PF and FAR indices should not be viewed as being fully reliable statistical measures of forecasting performance; on the other hand, for a given database, they do provide meaningful comparisons to be drawn among different icing algorithms (Brown et al., 1994).

The Hanssen and Kuipers (1965) discriminant V, defined as the hit rate minus the false alarm rate (PF - FAR), provides an acceptable and unbiased measure of forecast accuracy for scientific purposes. The score ranges from -1 to 1; -1 implies perfectly wrong forecasts, 0, random performance (PF = FAR), and 1, perfect skill. Since the V discriminant does not depend on the sample relative frequency of the predictand, forecast successes and failures are given equal weight. In general, the greater the positive score, the greater the likelihood for high hit rates to be associated with low false alarm rates. Hanssen and Kuipers' contingency table formulation of variance will be computed in order to assess whether or not differences between icing predictors are statistically significant. Given values of V for two predictors, the difference between them will be considered to be 'highly statistically significant' provided the skill score difference is greater than the standard deviation in the difference times a confidence factor set at a 0.01 level of significance (i.e., 99% confidence level). When used in this report, a 'statistically significant ' difference corresponds to a 0.05 level of significance (95% confidence level).

Statistical indices are computed for five distinct layers (LOW to HIGH) as well as for a columnal layer (COLUMN) which

encompasses all icing data from the LOW through HIGH layers. For NOGAPS data, discrete yes/no icing forecasts for a particular algorithm are based on only one model-derived value for the LOW-MID, MIDDLE, MID-HIGH and HIGH layers (corresponding to model 850, 700, 500 and 400 mb pressure levels, respectively) and two values for the LOW layer (derived from 1000 and 925 mb data). Only pilot reports which specify icing intensity are used for verification of binary yes/no icing forecasts. All four icing algorithms provide a specific 'no icing' forecast category. A designation of 'no icing' for a COLUMN layer (PIREP, model or RAOB) is only made provided the surface (i.e., lowest possible) icing layer (determined using terrain height) and all layers above explicitly have 'no icing' category values. Both the NAWAU and TESS algorithms are utilized twice. The predictor NAWAU(1) provides a positive icing forecast if either category 1 or 2 is forecast. NAWAU(2) is positive only if category 2 icing is forecast within the specified layer; for this predictor, category 1 icing is considered a negative forecast. The TESS(YALL) predictor yields a positive icing layer forecast if the probability of icing is non-zero (i.e., 10, 20, 50 or 100%). Much more restrictive, TESS(Y100) provides a positive icing forecast only if the algorithm forecasts a 100% probability of icing at any one level within the specified layer. TESS(Y100) icing forecasts of 10, 20 and 50% probabilities are all considered 'no icing' forecasts, and are grouped with 0% probability forecasts.

Verification of icing intensity and type is possible using both the AIRF and TESS algorithms. (Note: the paucity of freezing precipitation reports precludes any verification of RAP's ZR

and ZL icing categories.) For any specified icing layer (LOW to HIGH, or COLUMN), the percent frequency of agreement between reported and forecast (model or RAOB based) icing type and intensity as a function of reported type and intensity will be determined. These statistics, based on only positive (i.e., yes) icing reports, include the frequency of agreement in type, intensity, both type and intensity and, neither type nor intensity. Type comparisons exclude reports of freezing precipitation (ZR and ZL), since these icing types are not forecast by either algorithm. Because the AIRF and TESS algorithms forecast fewer intensity categories than were reported, the reported intensities are consolidated prior to comparison. Pilot report categories TRC-LGT and LGT-MDT were combined with the next higher category (LGT and MDT, respectively) prior to comparison with either icing algorithm. Reports of MDT-HVY, MDT-SVR and SVR were consolidated into a single SVR category for TESS comparisons. Since the AIRF algorithm does not forecast severe icing, any reported icing intensity between LGT-MDT and SVR was assigned a MDT intensity prior to comparison with AIRF forecasts. Incompatibility between observed and TESS(YALL) forecast icing type is resolved by excluding from comparisons all observed/forecast data pairs having a forecast type 'induction' and intensity 'unknown.' A similar incompatibility problem does not occur for the TESS(Y100) predictor since all forecasts of induction icing have been previously set to a 'no icing' category prior to determination of icing type and intensity for a particular layer. Given PIREP, model or radiosonde icing types and intensities for individual (LOW

through HIGH) layers, the composite COLUMN icing type is determined in an analogous procedure as that for any single PIREP or radiosonde icing layer with multiple icing data (previously discussed in sections 3.2.3 and 3.3).

A comparison of PIREPs used in this study with previous observational studies (Table 7) indicates that the most prevalent icing intensity and type is light rime. In order to assess the performance of the AIRF and TESS algorithms, icing intensity and type forecasts will be compared to the predictor 'LGT RME.' This predictor always assigns the icing intensity and type 'light rime' to any positive icing forecast. Specifically, given any particular AIRF or TESS positive icing forecast (intensity trace and above; type rime, mixed or clear), the existing icing intensity and type is converted to 'light rime' prior to comparison with the corresponding pilot report. The predictor 'LGT RME' is applied separately to icing determinations of each of the three predictors (AIRF, TESS(YALL) and TESS(Y100)), for each of the six icing layers (LOW to HIGH, COLUMN) and four categories of agreement statistics.

5. COMPARISON RESULTS

5.1 Model and PIREPs

Overall verification statistics from PIREP-based verification of icing algorithms using NOGAPS 2.5° analysis, 12 and 24 hr model data are presented in Table 9. For each of six icing predictors, the prefigurance and false alarm rate indices, and the V discriminant, are given for selected icing layers (LOW to HIGH, COLUMN). Statistical differences in skill scores, at a 0.01 level of significance, are indicated by numbers appended to

Table 9a. For selected icing layers (LOW to HIGH, COLUMN), verification statistics (PF, FAR, V) from PIREP-based verification of icing predictors with Tau = 0 hr NOGAPS 2.5° model data.

(a)

TAU = 0 HR	LAYER					
	LOW	LOW-MID	MIDDLE	MID-HIGH	HIGH	COLUMN
NO.PIREPS(Y/N)	71/126	266/373	549/515	237/497	29/474	950/106
			PREFIGURANCE (PF)			
AIRF	0.183	0.455	0.288	0.321	0.310	0.564
NAWAU(1)	0.493	0.778	0.701	0.473	0.0	0.868
NAWAU(2)	0.451	0.677	0.517	0.135	0.0	0.688
RAP	0.493	0.767	0.689	0.342	0.0	0.855
TESS(YALL)	0.0	0.752	0.521	0.582	0.448	0.817
TESS(Y100)	0.0	0.526	0.344	0.333	0.207	0.560
			FALSE ALARM RATE (FAR)			
AIRF	0.063	0.107	0.150	0.095	0.023	0.179
NAWAU(1)	0.143	0.214	0.363	0.089	0.002	0.358
NAWAU(2)	0.143	0.169	0.243	0.018	0.0	0.255
RAP	0.135	0.212	0.355	0.068	0.0	0.358
TESS(YALL)	0.008	0.268	0.317	0.189	0.070	0.330
TESS(Y100)	0.0	0.131	0.179	0.056	0.017	0.160
			HK DISCRIMINANT (V)			
AIRF	0.120	0.348	0.138	0.226	0.287	0.385
NAWAU(1)	0.350^{2}	0.564^{2}	0.338^{3}	0.384^{2}	-0.002	0.510
NAWAU(2)	0.308^{2}	0.508^{1}	0.275^{1}	0.117	0.0	0.434
RAP	0.358^{2}	0.555^{2}	0.333^{3}	0.273^{1}	0.0	0.496
TESS(YALL)	-0.008	0.484^{1}	0.204	0.393^{2}	0.379^{3}	0.487
TESS(Y100)	0.0	0.395	0.166	0.277^{1}	0.190	0.400

____x number of lower score predictors difference is at 0.01 level of significance

Table 9b. Same as a), except with Tau = 12 hr NOGAPS data.

(b)

TAU = 12 HR	LAYER					
	LOW	LOW-MID	MIDDLE	MID-HIGH	HIGH	COLUMN
NO.PIREPS(Y/N)	68/122	265/358	568/504	248/493	31/466	969/101
			PREFIGURANCE (PF)			
AIRF	0.235	0.291	0.234	0.387	0.323	0.549
NAWAU(1)	0.338	0.717	0.683	0.492	0.065	0.845
NAWAU(2)	0.338	0.558	0.537	0.165	0.0	0.651
RAP	0.338	0.713	0.678	0.375	0.0	0.829
TESS(YALL)	0.0	0.766	0.533	0.653	0.516	0.836
TESS(Y100)	0.0	0.415	0.375	0.375	0.129	0.558
			FALSE ALARM RATE (FAR)			
AIRF	0.090	0.089	0.111	0.105	0.026	0.188
NAWAU(1)	0.131	0.162	0.343	0.120	0.002	0.356
NAWAU(2)	0.115	0.106	0.212	0.018	0.0	0.198
RAP	0.131	0.159	0.321	0.067	0.002	0.347
TESS(YALL)	0.008	0.196	0.288	0.221	0.099	0.287
TESS(Y100)	0.0	0.092	0.153	0.071	0.022	0.109
			HK DISCRIMINANT (V)			
AIRF	0.145	0.201	0.123	0.282	0.297	0.361
NAWAU(1)	0.207	0.555^{2}	0.340^{2}	0.372^{1}	0.062	0.489
NAWAU(2)	0.223	0.452^{1}	0.325^{1}	0.147	0.0	0.453
RAP	0.207	0.554^{2}	0.356^{3}	0.308^{1}	-0.002	0.482
TESS(YALL)	-0.008	0.571^{2}	0.246^{1}	0.432^{2}	0.417^{3}	0.549^{1}
TESS(Y100)	0.0	0.323	0.222	0.304^{1}	0.107	0.449

____x number of lower score predictors difference is at 0.01 level of significance

Table 9c. Same as a), except with Tau = 24 hr NOGAPS data.

(c)

TAU = 24 HR	LAYER					
	LOW	LOW-MID	MIDDLE	MID-HIGH	HIGH	COLUMN
NO.PIREPS(Y/N)	65/119	255/354	524/493	232/477	29/456	905/100
			PREFIGURANCE (PF)			
AIRF	0.200	0.310	0.229	0.405	0.379	0.549
NAWAU(1)	0.431	0.694	0.626	0.522	0.034	0.833
NAWAU(2)	0.431	0.561	0.490	0.190	0.0	0.649
RAP	0.431	0.678	0.624	0.379	0.0	0.800
TESS(YALL)	0.046	0.737	0.477	0.716	0.655	0.831
TESS(Y100)	0.046	0.427	0.294	0.409	0.172	0.554
			FALSE ALARM RATE (FAR)			
AIRF	0.050	0.076	0.085	0.096	0.044	0.140
NAWAU(1)	0.151	0.161	0.306	0.111	0.002	0.360
NAWAU(2)	0.134	0.102	0.162	0.029	0.0	0.180
RAP	0.160	0.158	0.282	0.084	0.0	0.340
TESS(YALL)	0.008	0.195	0.245	0.212	0.101	0.250
TESS(Y100)	0.0	0.102	0.122	0.086	0.022	0.130
			HK DISCRIMINANT (V)			
AIRF	0.150	0.234	0.144	0.309[1]	0.335	0.409
NAWAU(1)	0.280	0.533[2]	0.320[2]	0.410[1]	0.032	0.473
NAWAU(2)	0.296	0.459[1]	0.328[2]	0.160	0.0	0.469
RAP	0.271	0.520[2]	0.342[3]	0.295	0.0	0.460
TESS(YALL)	0.038	0.542[2]	0.232	0.504[4]	0.554[4]	0.581
TESS(Y100)	0.046	0.326	0.172	0.324[1]	0.150	0.424

____x number of lower score predictors difference is at 0.01 level of significance

individual V scores. Specifically, for a given icing predictor, an appended number indicates how many of the lowest V scores (within the same icing level) are highly statistically different from the score of the selected predictor. As an example, consider the MIDDLE column statistics for Tau = 0 hr (Table 9a); here, both the NAWAU(1) and RAP skill scores are highly statistically different from the lowest three scores (i.e., those for AIRF, TESS(YALL) and TESS(Y100)), and the NAWAU(2) skill score is highly statistically different from the lowest V score (corresponding to the AIRF predictor).

As indicated in Table 9, the ratio of yes/no icing reports used for verification is most even for the MIDDLE column. Verification of icing forecasts within the HIGH layer is strongly skewed toward negative icing occurrences. On the other hand, over 90% of verifying PIREPs for the cumulative COLUMN layer correspond to positive icing. The number of negative icing PIREPs used for verification decreases slightly with forecast length within all icing layers. The number of positive icing PIREPs also decreases over the full forecast interval for the lowest two layers; however, for the upper three layers and the COLUMN layer, the number of positive icing reports is greatest at Tau = 12 hr, least at Tau = 24 hr.

Concentrating first on analysis (Tau = 0 hr) verification statistics, one observes that the NAWAU(1) and RAP icing predictors (both refinements of the Schultz- Politovich icing algorithm) have the best capability in diagnosing icing occurrence for the LOW to MIDDLE, and COLUMN, layers. High PF values, which

exceed 0.75 and 0.85 for the LOW-MID and COLUMN layers, respectively, are likely abetted by these two predictors' relatively low RH thresholds for cloud icing (Note: RAP's dominant icing categories are 'stable' and 'unstable'.) At the same time, these "generous" moisture thresholds result in FAR indices for these two predictors among the highest of all predictors for the LOW to MIDDLE and COLUMN layers. NAWAU(1) and RAP forecast capability decreases noticeably from the MIDDLE to MID-HIGH layers, then becomes nonexistent at the HIGH layer, as model temperatures fall below algorithm icing threshold limits. With more restrictive (i.e., higher) moisture thresholds for icing prediction, the NAWAU(2) predictor has somewhat lower PF and FAR indices than those for NAWAU(1). A widening gap in NAWAU(1) and NAWAU(2) forecast capability occurs from the LOW to MID-HIGH layers; at the MID-HIGH layer, model RH values are mostly well below the NAWAU(2) threshold limit of 75%. For the LOW layer, both TESS predictors indicate no forecast capability. Since all other algorithms forecast positive icing events within the LOW layer, this result likely has much more to do with the way the TESS algorithm computes icing than with actual T and Td icing thresholds. Under usual circumstances (viz., no saturation at the lowest model level), the TESS algorithm requires at least two levels to define a cloud icing layer. Therefore, unless model terrain heights corresponding to positive icing reports within the LOW layer are at or below 1000 mb, positive icing will generally not be forecast. Since all but one of combined TESS(YALL) and TESS(Y100) LOW layer icing forecasts are negative, most of these TESS forecasts likely were derived from model profile data

commencing at 925 mb. From the LOW to LOW-MID layer, the TESS(YALL) PF value jumps from zero to about 0.75; this value is comparable to those for NAWAU(1) and RAP. With its ability to forecast icing at lower temperatures than other predictors, TESS(YALL) has the highest PF and FAR scores at both the MID-HIGH and HIGH layers. Since TESS(Y100) limits icing to a more restrictive temperature range ($0 \geq T \geq -22\ ^{o}C$), its PF and FAR statistics are generally considerably lower than those for TESS(YALL). Limited by quite stringent moisture thresholds, the AIRF predictor's forecast capability ranks below most of the other predictors except at the HIGH layer, where its PF score ranks second to that for TESS(YALL). For the LOW-MID and MIDDLE layers, FAR indices for AIRF are the lowest (i.e., best) of any predictor.

In terms of skill score (V), the NAWAU(1) and RAP predictors are best for the lowest three layers (LOW to MIDDLE) and the NAWAU(2) predictor third best (Table 9a). For the LOW-MID and MIDDLE layers, the V differences between both NAWAU(1) and RAP, and either AIRF or TESS(Y100), are highly statistically significant. At the MID-HIGH layer, NAWAU(1) and TESS(YALL) exhibit the best skill, and NAWAU(2) the least. The TESS(YALL) predictor is the most skillful within the HIGH layer; differences in V scores between this predictor and NAWAU(1), NAWAU(2) and RAP are all highly statistically significant. Due to virtually nil icing forecast capability, HK discriminant values are zero or near zero for TESS at the LOW layer and, NAWAU and RAP at the HIGH layer. For the COLUMN layer, skill scores for NAWAU(1), RAP and TESS(YALL) are the highest (all near 0.5); however, no

highly statistically significant differences in V are found among any of the six icing predictors. A comparison of Tau = 0 hr LOW to HIGH layer skill scores indicates that, for all icing algorithms, the highest score (i.e., best overall forecasting skill) occurred at the LOW-MID icing layer. Interestingly, for this layer, the average skill score of all six predictors exceeds that for the COLUMN layer by about 0.25.

A comparison of 12 and 24 hr icing verification statistics (Tables 9b and c) with 0 hr (Table 9a) results indicate that, while some variations in statistical rankings (PF, FAR and V) do occur, the significant forecasting relationships found among the various icing algorithms when using model analysis data (detailed above) generally hold for model-derived icing forecasts at the longer 12 and 24 hr lengths. In general, for any selected icing layer, the predictor with the best PF statistical value (or the predictor with the best V score) is the same for 0, 12 and 24 hr forecasts. An exception to this result only occurs at the LOW-MID layer, where the NAWAU(1) and RAP prefigurance values are slightly higher than TESS(YALL) at Tau = 0 hr, but are somewhat lower than TESS(YALL) at both Tau = 12 and 24 hr. At the MID-HIGH layer, all six predictors indicate an increase in forecast capability over the 0 to 24 hr forecast interval; a similar upward trend in PF values with lead time occurs for TESS(YALL) and AIRF at the HIGH layer. This result relates directly with the steadily increasing positive model bias in moisture at mid tropospheric levels (see Table 8); given higher model moisture over time, more positive icing forecasts should occur as algorithmic moisture thresholds are increasingly surpassed. Statistical tests

indicate that differences in V scores as a function of forecast lead time are generally not significant; the only exception to this result occurs for the AIRF predictor, whose V score difference between Tau = 0 and 12 hr at the LOW-MID layer is highly statistically significant. The NAWAU(1) predictor's three 'best skill' rankings at Tau = 0 hr (for the LOW-MID, MIDDLE and COLUMN layers) are not repeated at either Tau = 12 or 24 hr, being replaced by either TESS(YALL) (LOW-MID and COLUMN layers) or RAP (MIDDLE layer). For any given layer, the top skill score among predictors is the same at both Tau = 12 and 24 hr; the TESS(YALL) predictor has the highest V scores at all layers except LOW and MIDDLE, where the NAWAU(2) and RAP predictors, respectively, score best. With lower NAWAU and RAP LOW layer skill scores at 12 and 24 hr (compared to those at 00 hr), differences in V between these predictors and TESS are not highly statistically significant at either 12 or 24 hr. The superiority of TESS(YALL) at upper levels (MID-HIGH and HIGH) is most evident at Tau = 24 hr, where differences in V between this predictor and four others are highly statistically significant.

In order to further examine the effect of forecast length on model-derived icing forecasts, performance statistics based on best overall statistical value over the 0 to 24 hr forecast interval were computed. For a given statistical measure (PF, FAR, V), Table 10 gives the number of predictors with the best statistical value at Tau = 0 hr, at Tau = 12 hr and at Tau = 24 hr, as a function of icing layer. For example, consider the skill score discriminant for the COLUMN layer; here, 2 of 6 icing

Table 10. For a given statistical measure (PF, FAR, V), the number of predictors with the best statistical value at Tau = 0, at Tau = 12 and at Tau = 24 hr, for selected icing layers.

	PREFIGURANCE			FALSE ALARM RATE			HK DISCRIMINANT		
TAU(HR)	0	12	24	0	12	24	0	12	24
LAYER									
LOW	3	1	2	0	3	1	3	0	3
LOW-MID	5	1	0	0	1	5	5	1	0
MIDDLE	3	3	0	0	0	6	0	4	2
MID-HIGH	0	0	6	4	1	0	0	1	5
HIGH	1	1	2	3	0	0	1	1	2
COLUMN	5	1	0	0	2	4	2	1	3
TOTAL	17	7	10	7	7	16	11	8	15

Table 11. For a given statistical measure (PF, FAR, V), the number of icing layers with the best statistical value at Tau = 0, at Tau = 12 and at Tau = 24 hr, for selected icing predictors.

	PREFIGURANCE			FALSE ALARM RATE			HK DISCRIMINANT		
TAU(HR)	0	12	24	0	12	24	0	12	24
PREDICTOR									
AIRF	3	1	2	2	0	4	1	0	5
NAWAU(1)	4	1	1	1	2	2	3	2	1
NAWAU(2)	3	1	1	0	1	3	2	0	3
RAP	4	0	1	0	2	3	3	2	0
TESS(YALL)	0	3	3	2	0	3	0	2	4
TESS(Y100)	3	1	2	2	2	1	2	2	2
TOTAL	17	7	10	7	7	16	11	8	15

predictors had their highest V score at Tau = 0 hr, one at Tau = 12 hr, and 3 at Tau = 24 hr. In some cases, a particular statistical measure might not have a 'best' score at any one particular forecast length; when this occurred, that particular predictor was not included statistically and the Tau = 0, 12 and 24 hr sum of best predictor scores for the given layer was less than six (ex., HK discriminant, HIGH layer; 0, 12, and 24 hr sum is only 4).

Table 10 numerical totals (of all six icing layers) indicate that collectively, while this study's icing algorithms had best forecast capability (i.e., highest PF scores) at Tau = 0 hr, they had fewest false alarms and overall best skill (V scores) at the 24 hr forecast length. In a collective sense, forecasting performance at Tau = 12 hr was the poorest of the three forecast lead times. Within the LOW-MID and COLUMN layers, most predictors had best PF values using analysis (Tau = 0 hr) data; on the contrary, at the MID-HIGH level, all predictors had best PF scores when using 24 hr forecast data. While FAR scores for most predictors are best at Tau = 24 hr for the LOW-MID and MIDDLE layers, such scores for the highest two layers (MID-HIGH and HIGH) are best at Tau = 0 hr. Five of the six predictors exhibit best forecasting skill (i.e., highest V score) within the LOW-MID (MID-HIGH) layer at Tau = 0 hr (Tau = 24 hr). This particular result correlates well with analogous PF statistics, but poorly with those for FAR scores, suggesting that the skill score discriminants (PF - FAR) of study icing algorithms are more sensitive to high PF values than low false alarm rates. Although 12 hr icing predictions were overall less skillful than

those based on model 0 or 24 hr data, four of six predictors had highest V scores for the MIDDLE layer at Tau = 12 hr.

Overall forecasting performance may also be examined in terms of individual predictors. For a given statistical measure (PF, FAR, V), Table 11 gives the number of icing layers (LOW to HIGH, COLUMN) with best statistical value at Tau = 0, 12 and 24 hr, for selected icing predictors. For both NAWAU predictors and RAP, highest PF values mostly occur when predictions are based on model analysis data while lowest FAR scores generally occur when icing forecasts are based on NOGAPS 12 and 24 hr data. TESS(YALL) forecast capability is better at the longer (12 and 24 hr) forecast lengths. For the majority of icing layers, both the AIRF and TESS(YALL) predictors exhibit highest skill at Tau = 24 hr. Based on Table 11 data, the NAWAU(1) and RAP predictors show a steady decline in forecasting skill (V) with lead time, while the opposite occurs for TESS(YALL).

For a first examination of the AIRF and TESS algorithms' ability to differentiate icing type and intensity, the observed number of instances of selected icing types and intensities compared with number forecast, based on COLUMN layer data at Tau = 0 hr, is presented in Figure 4. Results indicate that both icing algorithms considerably overforecast 'no icing' conditions. Specifically, the TESS(YALL) and AIRF forecast/observed ratios of 'no icing' conditions are between 4:1 and 5:1, while the number of TESS(Y100) forecasts of 'no icing' exceeds number observed by about a 2.5 to 1 ratio. Of positive icing reports, the vast majority are of intensities light and moderate (Figure 4a).

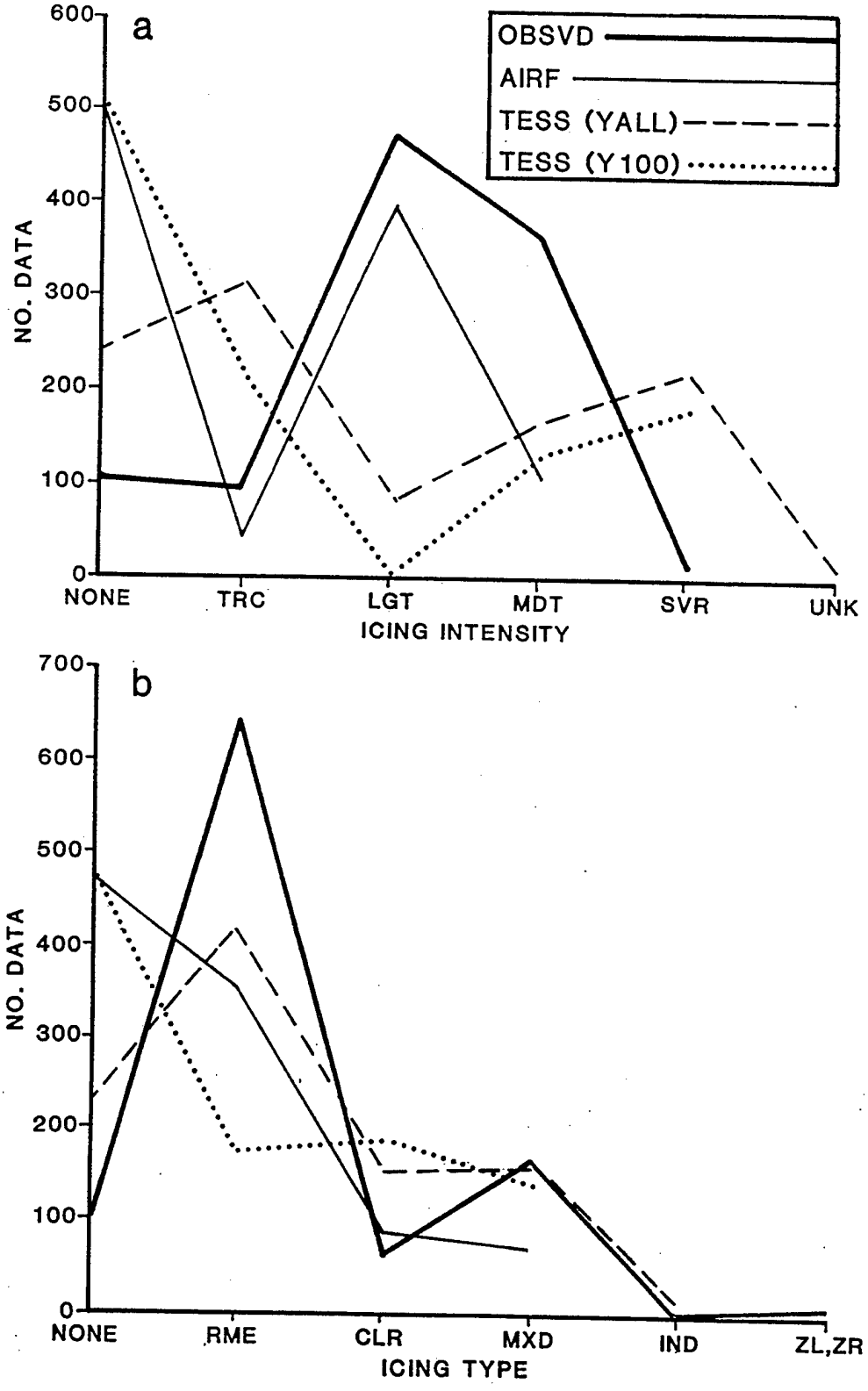

Figure 4. a. The observed number of selected icing intensities compared with number forecast by the AIRF, TESS(YALL) and TESS(Y100) predictors, based on COLUMN layer data at Tau = 0 hr. b. Same as a., except for selected icing types.

While the AIRF algorithm only somewhat underforecasts occurrences of both TRC and LGT icing intensities, it considerably underforecasts (along with both TESS predictors) occurrences of MDT intensity icing. Both TESS predictors, most especially TESS(Y100), greatly underpredict the most observed icing intensity - light. On the other hand, TESS(YALL) and TESS(Y100) both yield many more forecasts of severe icing conditions than observed; a similar result holds for the TRC icing category. Figure 4b indicates that rime icing is by far the most dominant icing type, being reported four times as often as mixed and about 10 times as often as CLR. Rime icing is underforecast by all three predictors, most especially TESS(Y100). Both TESS predictors overforecast clear icing conditions; however, the number of mixed icing reports is similar to the number forecast by each of the TESS predictors. Finally, while the AIRF algorithm forecasts less than half (42%) as many mixed icing conditions as actually observed, this predictor's number of clear icing forecasts is somewhat greater (by 24) than the observed number of such icing events.

To further examine the analysis (Tau = 0 hr) COLUMN layer dataset of Figure 4, percent frequencies of agreement between observed and forecast icing type and intensity are determined. These percentages are computed both as a function of observed, and forecast, type and intensity (Tables 12a and b, respectively). TESS(YALL) compilations of Table 12a do not include 'induction' icing forecasts, while Table 12b tabulations exclude reported ZR,ZL icing conditions. For this particular dataset, TESS(YALL) correctly forecasts almost half of all reported rime

Table 12. a) For the AIRF, TESS(YALL) and TESS(Y100) predictors, percent frequencies of agreement between observed and forecast icing type and intensity for the COLUMN layer at Tau = 0 hr, as a function of observed type and intensity. b) Same as a), except as a function of forecast type and intensity.

PREDICTOR	ICING TYPE			ICING INTENSITY			
	RME	CLR	MXD	TRC	LGT	MDT	SVR
a)							
AIRF	238/644	8/62	13/172	6/97	178/476	42/377	———
	37.0%	12.9%	7.6%	6.2%	37.4%	11.1%	———
TESS(YALL)	305/634	16/62	36/169	35/97	27/464	56/356	3/18
	48.1%	25.8%	21.3%	36.1%	5.8%	15.7%	16.7%
TESS(Y100)	130/644	20/62	34/172	27/97	4/476	46/359	3/18
	20.2%	32.3%	19.8%	27.8%	0.8%	12.8%	16.7%
b)							
AIRF	238/351	8/86	13/72	6/49	178/398	42/108	———
	67.0%	9.3%	18.1%	12.2%	44.7%	38.9%	———
TESS(YALL)	305/422	16/151	36/167	35/313	27/86	56/173	3/224
	72.3%	10.6%	21.6%	11.2%	31.4%	32.4%	1.3%
TESS(Y100)	130/177	20/187	34/141	27/227	4/4	46/135	3/183
	73.4%	10.7%	24.1%	11.9%	100.0%	34.1%	1.6%

icing events, AIRF 3 out of 8 such occurrences, and TESS(Y100), only 1 of 5. Both TESS predictors outperform AIRF in forecasting observed CLR and MXD icing events, with percent frequencies of agreement for TESS(YALL) and TESS(Y100) between 2 and 3 times higher than AIRF. In regards to icing intensity, the AIRF predictor is far better than either TESS predictor in assessing reported light icing intensities; even so, the AIRF algorithm only correctly forecasts 3 out of every 8 such events. Both TESS predictors forecast TRC icing events considerably better than AIRF; nonetheless, the best predictor (TESS(YALL)) is only able to correctly assess about 3 of 8 trace icing occurrences. For moderate and severe icing, all three predictors provide similarly low forecasting capability, with percent frequencies of agreement between 11 and 17 %. Table 12b results indicate that, in spite of noticeable differences among predictors in the number of icing type forecasts, percent frequencies of agreement for any selected icing type are similar for all three predictors. Forecasts of RME icing are most accurate, with verification rates about 70% for each predictor. Both CLR and MXD icing types are significantly overforecast by each predictor, with only about 10% (20%) of CLR (MXD) icing forecasts actually verifying. For TRC and LGT intensities, large differences are noted among predictors in the number of issued forecasts; for example, at the LGT intensity category, almost 400 forecasts were issued by the AIRF algorithm, only 4 by TESS(Y100). For either the TRC or MDT icing category, percent frequencies of agreement are similar for all three icing predictors; however, percentages for MDT icing

are about 3 times higher than those for the LGT icing category. The AIRF predictor's best verification of icing occurs for the highly reported LGT category, with almost a 45% frequency of agreement; interestingly, all 4 TESS(Y100) forecasts of light icing verified. Overforecasting of severe icing conditions is a serious problem for both TESS predictors, with many forecasts (~ 200) and few verifications (3).

Tables 13a-c present percent frequencies of agreement between reported and forecast (AIRF, TESS(YALL), TESS(Y100)) icing type and intensity, for 6 icing layers (LOW to HIGH, COLUMN) and 3 forecast lengths (Tau = 0, 12, 24 hr), as a function of observed type and intensity. Within these tables, the first value in the "No. Data" column for a given layer and statistic corresponds to both the AIRF and TESS(Y100) predictors, while the second value (in parenthesis) corresponds to the TESS(YALL) predictor. In general, the number of data for the TESS(YALL) predictor (at the LOW-MID, MIDDLE and COLUMN layers) are reduced somewhat with respect to the other two predictors, since 'induction' type icing forecasts are not included in the statistics. For a given layer, forecast tau and agreement statistic, the percentage in parenthesis after any of the three icing predictors, corresponds to the percent frequency of agreement for the 'LGT RME' predictor based on that particular observation/predictor dataset.

Results presented in Table 13 indicate that, regardless of the icing layer or forecast length, or which of the four icing agreement classifications is selected, percent frequencies of agreement for TESS(Y100) are always the worst (or tied for the

Table 13a. For the AIRF, TESS(YALL) and TESS(Y100) predictors, percent frequencies of agreement between reported and forecast icing type and intensity based on selected icing agreement classifications, for selected icing layers (LOW to HIGH, COLUMN) at Tau = 0 hr. Percentages in parentheses correspond to 'LGT RME' predictor (see text for further information).

(a)

TAU = 0 HR		PREDICTOR			
	NO.DATA	AIRF		TESS(YALL)	TESS(Y100)
		TYPES AGREE			
LAYER					
LOW	67 (67)	9.0%	(9.0%)	0.0% (0.0%)	0.0% (0.0%)
LOW-MID	241(220)	13.3%	(32.4%)	37.7% (49.5%)	27.8% (34.9%)
MIDDLE	516(499)	16.9%	(20.5%)	20.8% (37.7%)	14.3% (24.2%)
MID-HIGH	217(217)	24.0%	(25.3%)	25.3% (44.7%)	6.9% (23.0%)
HIGH	27 (27)	25.9%	(25.9%)	18.5% (37.0%)	0.0% (18.5%)
COLUMN	878(865)	29.5%	(40.0%)	41.3% (58.6%)	21.0% (39.0%)
		INTENSITIES AGREE			
LAYER					
LOW	71 (71)	7.0%	(8.5%)	0.0% (0.0%)	0.0% (0.0%)
LOW-MID	266(241)	21.8%	(22.2%)	9.1% (35.3%)	7.1% (26.7%)
MIDDLE	549(532)	10.6%	(14.4%)	9.0% (23.1%)	7.5% (16.9%)
MID-HIGH	237(237)	13.5%	(16.5%)	3.4% (29.1%)	1.7% (14.8%)
HIGH	29 (29)	13.8%	(13.8%)	3.4% (24.1%)	3.4% (10.3%)
COLUMN	950(935)	23.8%	(26.1%)	12.9% (39.4%)	8.4% (26.5%)
		BOTH TYPE AND INTENSITY AGREE			
LAYER					
LOW	66 (66)	1.5%	(3.0%)	0.0% (0.0%)	0.0% (0.0%)
LOW-MID	236(215)	7.2%	(13.1%)	4.2% (22.3%)	2.5% (15.3%)
MIDDLE	512(495)	5.7%	(10.7%)	4.0% (17.4%)	3.9% (11.9%)
MID-HIGH	216(216)	8.3%	(12.0%)	2.3% (22.7%)	0.5% (8.8%)
HIGH	27 (27)	11.1%	(11.1%)	0.0% (22.2%)	0.0% (7.4%)
COLUMN	869(856)	12.3%	(18.1%)	8.4% (29.3%)	4.4% (18.1%)
		NEITHER TYPE NOR INTENSITY AGREE			
LAYER					
LOW	66 (66)	84.8%	(84.8%)	100.% (100.%)	100.% (100.%)
LOW-MID	236(215)	71.2%	(59.7%)	58.6% (38.1%)	69.1% (54.7%)
MIDDLE	512(495)	77.9%	(75.4%)	74.1% (56.2%)	81.6% (70.3%)
MID-HIGH	216(216)	70.8%	(70.4%)	73.1% (48.6%)	91.7% (72.2%)
HIGH	27 (27)	70.4%	(70.4%)	77.8% (59.3%)	96.3% (77.8%)
COLUMN	869(856)	58.5%	(52.0%)	53.9% (30.4%)	74.7% (52.6%)

Table 13b. Same as a), except at Tau = 12 hr.

(b)

TAU = 12 HR	NO.DATA	AIRF	PREDICTOR TESS(YALL)	TESS(Y100)
		TYPES AGREE		
LAYER				
LOW	64 (64)	14.1% (14.1%)	0.0% (0.0%)	0.0% (0.0%)
LOW-MID	243(214)	13.2% (20.2%)	38.3% (51.9%)	20.6% (28.0%)
MIDDLE	534(515)	15.9% (17.6%)	19.2% (38.6%)	14.6% (27.3%)
MID-HIGH	227(227)	27.3% (30.4%)	22.5% (49.8%)	5.3% (28.6%)
HIGH	29 (29)	27.6% (27.6%)	31.0% (44.8%)	0.0% (13.8%)
COLUMN	898(882)	31.8% (39.1%)	39.3% (61.0%)	17.9% (40.2%)
		INTENSITIES AGREE		
LAYER				
LOW	68 (68)	8.8% (19.1%)	0.0% (0.0%)	0.0% (0.0%)
LOW-MID	265(233)	12.1% (14.3%)	9.0% (34.8%)	6.0% (21.1%)
MIDDLE	568(548)	9.3% (11.1%)	10.4% (23.5%)	8.5% (16.7%)
MID-HIGH	248(248)	15.3% (18.5%)	4.0% (33.1%)	1.2% (18.5%)
HIGH	31 (31)	16.1% (16.1%)	6.5% (29.0%)	3.2% (3.2%)
COLUMN	969(952)	22.8% (26.6%)	14.2% (40.0%)	7.6% (26.7%)
		BOTH TYPE AND INTENSITY AGREE		
LAYER				
LOW	63 (63)	4.8% (14.3%)	0.0% (0.0%)	0.0% (0.0%)
LOW-MID	238(209)	4.6% (8.8%)	3.8% (23.0%)	2.9% (13.0%)
MIDDLE	530(511)	6.0% (8.3%)	4.5% (17.2%)	4.2% (11.7%)
MID-HIGH	226(226)	11.9% (15.0%)	2.2% (27.0%)	0.0% (14.2%)
HIGH	29 (29)	13.8% (13.8%)	0.0% (27.6%)	0.0% (3.4%)
COLUMN	889(873)	13.7% (19.1%)	8.7% (30.1%)	3.6% (18.8%)
		NEITHER TYPE NOR INTENSITY AGREE		
LAYER				
LOW	63 (63)	81.0% (81.0%)	100.% (100.%)	100.% (100.%)
LOW-MID	238(209)	78.6% (73.5%)	57.4% (36.8%)	76.9% (64.3%)
MIDDLE	530(511)	80.8% (79.2%)	74.6% (54.6%)	80.6% (67.2%)
MID-HIGH	226(226)	69.0% (65.9%)	76.1% (43.4%)	93.4% (67.3%)
HIGH	29 (29)	69.0% (69.0%)	62.1% (51.7%)	96.6% (86.2%)
COLUMN	889(873)	58.8% (53.1%)	55.2% (28.3%)	77.7% (51.5%)

Table 13c. Same as a), except at Tau = 24 hr.

(c)

TAU = 24 HR	NO.DATA	AIRF	PREDICTOR TESS(YALL)	TESS(Y100)
		TYPES AGREE		
LAYER				
LOW	62 (62)	12.9% (12.9%)	1.6% (1.6%)	1.6% (1.6%)
LOW-MID	231(198)	14.7% (20.8%)	38.9% (48.9%)	24.2% (27.7%)
MIDDLE	491(468)	15.3% (17.5%)	19.0% (34.2%)	11.2% (21.6%)
MID-HIGH	212(212)	27.8% (30.7%)	25.0% (53.3%)	6.6% (29.2%)
HIGH	26 (26)	30.8% (34.6%)	46.2% (61.5%)	0.0% (15.4%)
COLUMN	835(819)	30.8% (38.7%)	38.9% (60.1%)	16.8% (38.9%)
		INTENSITIES AGREE		
LAYER				
LOW	65 (65)	12.3% (16.9%)	1.5% (3.1%)	1.5% (3.1%)
LOW-MID	255(218)	11.8% (16.5%)	10.1% (32.6%)	5.9% (21.6%)
MIDDLE	524(499)	8.8% (12.8%)	5.8% (21.0%)	4.4% (15.5%)
MID-HIGH	232(232)	19.4% (20.3%)	6.0% (36.2%)	2.6% (19.8%)
HIGH	29 (29)	13.8% (17.2%)	3.4% (34.5%)	3.4% (3.4%)
COLUMN	905(885)	23.5% (28.3%)	12.1% (40.1%)	5.3% (27.5%)
		BOTH TYPE AND INTENSITY AGREE		
LAYER				
LOW	61 (61)	9.8% (13.1%)	1.6% (0.0%)	1.6% (0.0%)
LOW-MID	228(195)	4.4% (11.4%)	5.1% (21.5%)	2.2% (13.2%)
MIDDLE	487(464)	6.2% (10.3%)	3.0% (15.7%)	2.3% (11.5%)
MID-HIGH	211(211)	13.7% (15.6%)	2.8% (28.0%)	0.0% (13.3%)
HIGH	26 (26)	11.5% (15.4%)	0.0% (34.6%)	0.0% (3.8%)
COLUMN	828(812)	13.3% (20.4%)	7.6% (29.6%)	2.3% (19.2%)
		NEITHER TYPE NOR INTENSITY AGREE		
LAYER				
LOW	61 (61)	83.6% (83.6%)	98.4% (95.1%)	98.4% (95.1%)
LOW-MID	228(195)	77.2% (72.8%)	56.4% (40.5%)	72.8% (64.5%)
MIDDLE	487(464)	81.5% (79.3%)	78.0% (59.7%)	86.4% (73.9%)
MID-HIGH	211(211)	66.4% (64.0%)	71.6% (37.4%)	90.5% (64.5%)
HIGH	26 (26)	65.4% (61.5%)	50.0% (34.6%)	96.2% (84.6%)
COLUMN	828(812)	58.2% (52.9%)	56.8% (28.3%)	80.2% (52.3%)

worst) among the three icing predictors. Regarding agreement in icing type, the TESS(YALL) predictor scored the highest percentages at the LOW-MID, MIDDLE and COLUMN layers at all 3 forecast lengths, while the AIRF predictor did the same at the LOW layer. In general, icing type percent frequencies of agreement for AIRF increase with layer height. For any individual layer, the most dramatic change over time in percent agreement of icing type occurs for TESS(YALL) at the HIGH layer, which rises from 18.5% at Tau = 0 hr to 46.2% at Tau = 24 hr. With only one exception (MIDDLE layer, Tau = 12 hr), the highest percent frequencies of agreement for icing intensity correspond to the AIRF predictor. For the most part, agreement percentages for intensity are less than corresponding values for icing type for all three predictors; in several instances, differences in percent frequency between type and intensity exceed 25% for TESS(YALL). As one might expect, frequency of agreement statistics requiring observed/forecast agreement in both type and intensity are the lowest among the 4 statistical classifications; except for the AIRF predictor at the COLUMN and higher layers, percent frequencies of agreement in 'both type and intensity' do not exceed 10%. Results for the classification 'neither type nor intensity agree' are similar to those for agreement in type, with the TESS(YALL) predictor yielding best (i.e., lowest) percent frequencies of agreement at the LOW-MID, MIDDLE and COLUMN layers at all 3 forecast lengths, and the AIRF predictor at the LOW and MID-HIGH layers. Additionally, for this particular classification, a noticeable improvement over time occurs for TESS(YALL) at the

HIGH layer, with a decrease in percentage from ~ 78% at Tau = 0 hr to 50% at Tau = 24 hr. This value of 50% represents the absolute lowest (i.e., best) percentage for the 'neither type nor intensity' agreement classification among all predictors, at all three forecast lengths. Finally, for any selected icing predictor, comparisons among analysis, 12 and 24 hr percent frequency of agreement statistics do not indicate any significant or definitive trends with time in the prediction of either icing type or intensity.

By far, the most intriguing results presented in Table 13 concern the comparisons among the AIRF and TESS predictors and the hypothetical 'LGT RME' predictor. For the data-rich layers (LOW-MID through MID-HIGH, COLUMN), percent frequencies of agreement for the 'LGT RME' predictor, for all forecast lengths and agreement classes, are always higher than corresponding percentages for any particular predictor (AIRF, TESS(YALL), TESS(Y100)) selected. Additionally, for the more data-poor layers (LOW and HIGH), percent frequencies of agreement for 'LGT RME' are equal or higher than those for AIRF or TESS predictors except for agreement in 'both type and intensity' at Tau = 24 hr, where both TESS predictors are marginally higher (at an insignificant 1.6%). The fact that a prediction of icing type and intensity, based on historical most likely categories (viz., light rime), is generally more accurate than what can be achieved by operational icing algorithms such as AIRF and TESS attests to the great difficulty such algorithms encounter when attempting to differentiate icing type and intensity.

Further examination of Table 13 results indicate that, for

the most part, differences in percent frequency of agreement between the 'LGT RME' predictor and TESS(YALL) are the largest while those between 'LGT RME' and the AIRF predictor are the smallest; in several instances (mostly at Tau = 24 hr), differences between 'LGT RME' and TESS(YALL) exceed 30%. For all layers except LOW, 'LGT RME' percent frequencies of agreement are always best when based on TESS(YALL) forecast/observed datasets; at the LOW layer, frequencies computed using AIRF datasets are best. For the COLUMN layer, TESS(YALL)-based 'LGT RME' percent frequencies are about 60% for type agreement, 40% for intensity agreement, and near 30% for agreement in 'both type and intensity' and 'neither type nor intensity'. When compared, these percent frequencies of agreement are considerably better than corresponding 'LGT RME' percentages based on either AIRF or TESS(Y100) datasets.

5.2 Model, Radiosondes and PIREPs

In this section, two main topics will be explored in depth; one, the skill level of model-based (NOGAPS) icing predictions compared to radiosonde-based predictions and two, the importance of vertical data resolution in icing prediction. For a given statistical measure (PF, FAR, V), Table 14 gives the cumulative number of icing predictors, summed over Tau = 0, 12 and 24 hr, with best statistical value based on NOGAPS (NG) and radiosonde (RAOB) data, at selected icing layers. Here, RAOB data includes both standard and significant levels, NOGAPS only standard levels plus 925 mb. Given a particular layer and statistic, the maximum number possible for either model- or RAOB-based predictions is 18

Table 14. For a given statistical measure (PF, FAR, V), the cumulative number of icing predictors, summed over Tau = 0, 12 and 24 hr, with best statistical value based on NOGAPS (NG) and radiosonde (RAOB) data, for selected icing layers.

	PREFIGURANCE		FALSE ALARM RATE		HK DISCRIMINANT	
DATA SOURCE	(NG)	(RAOB)	(NG)	(RAOB)	(NG)	(RAOB)
LAYER						
LOW	0	3	9	3	6	6
LOW-MID	0	16	18	0	2	16
MIDDLE	0	18	18	0	0	18
MID-HIGH	3	15	16	2	3	15
HIGH	1	5	7	7	6	8
COLUMN	3	15	8	4	6	12
TOTAL	7	72	76	16	23	75

Table 15. For a given statistical measure (PF, FAR, V), the cumulative number of icing layers, summed over Tau = 0, 12 and 24 hr, with best statistical value based on NOGAPS (NG) and radiosonde (RAOB) data, for selected icing predictors.

	PREFIGURANCE		FALSE ALARM RATE		HK DISCRIMINANT	
DATA SOURCE	(NG)	(RAOB)	(NG)	(RAOB)	(NG)	(RAOB)
PREDICTOR						
AIRF	0	13	13	4	3	15
NAWAU(1)	1	12	15	1	6	10
NAWAU(2)	0	12	10	1	0	13
RAP	0	11	15	1	4	12
TESS(YALL)	6	11	10	6	6	12
TESS(Y100)	0	13	13	3	4	13
TOTAL	7	72	76	16	23	75

(3 forecast taus x 6 predictors); since ties are not counted, NG + RAOB sums are often less than 18. Table 14 statistics are derived from tabular data presented as Appendix A. In general, the ratio of yes/no PIREPs upon which statistics are computed is highly skewed toward 'no' PIREPs at the LOW and HIGH layers, about even at the MIDDLE layer, and about 7:1 for the COLUMN layer. The total number of PIREPs used in model/radiosonde comparisons ranges from about 22 for the LOW layer to near 190 for the MIDDLE layer.

Table 14 prefigurance totals indicate a much stronger forecast capability when using radiosonde data than if predictions are based on NOGAPS data. For the lowest three layers, all PF index comparisons indicate higher (or equal) values for RAOB-based predictions. On the contrary, false alarm rates for NOGAPS-based predictions are decidedly superior except at the HIGH layer; for both the LOW-MID and MIDDLE layers, best FAR values for all 6 predictors always correspond to model-derived predictions. The observed behavior in PF and FAR statistics is likely related to differences in vertical resolution between model and radiosonde data; with more data levels, RAOB-based icing predictions are likely to have more hits and false alarms (i.e., higher PF, worst FAR). Overall Table 14 numerical totals for discriminant skill score show a decided advantage for RAOB-based predictions over those using model data (75 to 23). This advantage in forecasting skill manifests itself at the LOW-MID through MID-HIGH, and COLUMN, icing layers. At the LOW and HIGH layers, no decided advantage in forecasting skill is apparent for either model- or RAOB-based predictions. Skill score statistics

at these two layers are likely quite sensitive to the disproportionate ratios of yes/no icing reports, which include few positive icing occurrences.

For a given statistical measure, Table 15 gives the cumulative number of icing layers, summed over Tau = 0, 12 and 24 hr, with best statistical value based on NOGAPS and radiosonde data, for selected icing predictors. Similar to Table 14, the maximum possible sum (NG + RAOB) for a given predictor and statistic is 18 (3 forecast taus x 6 icing layers). For each predictor, forecast capability is decidedly superior when RAOB-based data is used for icing prediction. Interestingly, of the total 7 'best' PF index values for model-based (NG) forecasts, 6 of these correspond to the TESS(YALL) predictor. False alarm rates for each predictor are most often better when model data is utilized; for NAWAU(1) and RAP, 15 'best' FAR statistics correspond to NG data and only one to RAOB-based predictions. Regardless of which predictor is chosen, forecasting skill (V) is superior using radiosonde data; this superiority is most striking for NAWAU(2), with a RAOB to NG 'best' statistical value ratio of 13 to zero.

Statistical tests performed on skill score differences between model- and RAOB-based predictions (Appendix A) indicate that 8 such differences are highly significant while another 10 differences are significant at a 95% confidence level. Of these 18 differences, skill score values for RAOB-based forecasts were superior to those based on model data in all but one comparison. All significant and highly significant differences correspond to the layers LOW-MID, MIDDLE, MID-HIGH and COLUMN. Peculiarly, six

of the 8 highly significant differences correspond to the AIRF predictor.

Although designed primarily for model/radiosonde comparisons, Appendix A tables also permits algorithm forecast performance comparisons to be drawn among predictions based exclusively on model or radiosonde data. Examination of Appendix A radiosonde-based results indicate that, for the two layers with the most positive icing reports (COLUMN and MIDDLE), the TESS predictors (YALL and Y100) always rank lowest in forecast capability and skill (PF and V) among all 6 predictors, at all 3 forecast lengths. Of the 280 radiosondes utilized for Appendix A statistics, 80 (or 2 of every 7) had surface superadiabatic lapse rates. As a result of TESS algorithm processing of such conditions, negative icing was always specified at all layers of these radiosondes. On the contrary, for many of these same radiosondes, the other three algorithms (AIRF, NAWAU and RAP) often forecast positive icing to occur at one or more layers above such surface superadiabatic conditions. Thus, for layers reporting large number of positive icing events (viz., MIDDLE and COLUMN) , radiosonde-based TESS forecasting capability (PF) and skill (V) are noticeably less than for other icing predictors.

Collectively, results presented so far indicate that icing algorithms have better overall forecast capability and skill when based on full radiosonde data. This forecast advantage over model data could result from two key factors: one, higher vertical resolution and two, more accurate meteorological data. To explore this further, comparisons among model and radiosonde icing predictions were performed using reduced radiosonde data (designat-

ed RAOBR). Icing forecasts based on RAOBR data utilize radiosonde standard levels (1000, 850, 700, 500, 400 mb) plus the 925 mb surface if available. Thus, RAOBR data levels are essentially the same as those available with NOGAPS. If one assumes radiosonde (temperature, moisture and height) data to be accurate, then direct comparisons of RAOBR and NOGAPS-based icing predictions will provide a measure of the effect model data inaccuracies have on icing prediction. Appendix B provides full comparative statistics for model- and RAOBR-based 0, 12 and 24 hr forecasts.

For a given statistical value (PF, FAR, V), Table 16 gives the cumulative number of icing predictors (summed over all 3 forecast lengths) with best statistical value based on NOGAPS and reduced radiosonde (RAOBR) data, at selected icing layers. Cumulative totals for all 6 layers indicate somewhat better overall performance for RAOBR-based predictions, with numerical totals of all three statistical measures higher for RAOBR than NG. For the LOW, LOW-MID, MID-HIGH and COLUMN layers, best statistical value numbers for RAOBR-based predictions are higher than (or equal to) those for NOGAPS-based forecasts for all 3 statistical measures. Differences between RAOBR and NG sums are generally most noticeable at the MID-HIGH and COLUMN layers; at both layers, false alarm rates for icing predictions based on RAOBR data were superior to those for NG data 15 out of 18 times, with 3 ties. Table 16 comparison results for the MIDDLE and HIGH layers are both interesting. The MIDDLE layer, which contains both the highest number of data ($\sim$ 190) and the most even ratio

Table 16. For a given statistical measure (PF, FAR, V), the cumulative number of icing predictors, summed over Tau = 0, 12 and 24 hr, with best statistical value based on NOGAPS (NG) and reduced radiosonde (RAOBR) data, for selected icing layers.

	PREFIGURANCE		FALSE ALARM RATE		HK DISCRIMINANT	
DATA SOURCE	(NG)	(RAOBR)	(NG)	(RAOBR)	(NG)	(RAOBR)
LAYER						
LOW	0	3	0	0	0	3
LOW-MID	7	9	5	8	7	11
MIDDLE	9	8	11	5	9	9
MID-HIGH	5	12	0	15	4	14
HIGH	5	1	0	9	5	6
COLUMN	6	10	0	15	6	11
TOTAL	32	43	16	52	31	54

Table 17. For a given statistical measure (PF, FAR, V), the cumulative number of icing layers, summed over Tau = 0, 12 and 24 hr, with best statistical value based on NOGAPS (NG) and reduced radiosonde (RAOBR) data, for selected icing predictors.

	PREFIGURANCE		FALSE ALARM RATE		HK DISCRIMINANT	
DATA SOURCE	(NG)	(RAOBR)	(NG)	(RAOBR)	(NG)	(RAOBR)
PREDICTOR						
AIRF	8	8	2	9	8	10
NAWAU(1)	4	8	4	4	4	9
NAWAU(2)	2	9	3	6	2	10
RAP	2	8	3	9	2	10
TESS(YALL)	12	2	2	12	11	4
TESS(Y100)	4	8	2	12	4	11
TOTAL	32	43	16	52	31	54

of yes/no PIREPS (about 1:1), shows no advantage in forecasting capability or skill for either of the two data sources, but does have decidedly better false alarm rates for model-based forecasts. For the HIGH layer, Table 16 results indicate better PF statistics and worst false alarm rates for model-based icing predictions. These particular results could be influenced by the positive moisture bias in NOGAPS at the 400 mb (HIGH) layer, a condition conducive to enhanced positive icing prediction (and resultantly, more hits and false alarms). Collectively, Table 16 statistics (PF, FAR, V) show the best advantage of RAOBR-based predictions over those using model data at the MID-HIGH (500 mb) layer.

For any given icing predictor and statistical measure, Table 17 gives the cumulative number of icing layers with best statistical value based on NOGAPS and RAOBR-based data. Considering all three statistics, the predictors NAWAU (1 and 2), RAP and TESS(Y100) show somewhat better forecasting performance for reduced radiosonde data than NOGAPS. Forecasting capability and skill for the AIRF predictor are about the same for either model or RAOBR data, although false alarms for model-based forecasts are decidedly higher. By far, the most interesting Table 17 results concern TESS(YALL). Of all predictors, TESS(YALL) is the only one to show better forecast capability and skill for model-based predictions. Without TESS(YALL) numerical values, total sums for both PF and V would be considerably more skewed toward overall better performance with RAOBR data (i.e., PF, 20 to 41; V, 20 to 50). Careful examination of TESS(YALL) model- and RAOBR-based 0, 12 and 24 hr forecasts (Appendix B) reveals

that, for the LOW-MID, MIDDLE and COLUMN layers, all best statistical values for both PF and V correspond to model-based forecasts. In addition, at Tau = 24 hr, the MID-HIGH and HIGH layers also have best PF and V values for icing predictions based on NOGAPS data. While the enhanced TESS(YALL) model-based forecasting performance at and above the MIDDLE layer (especially at the longer 12 and 24 hr forecast lengths) might have been abetted by the positive NOGAPS moisture bias between 700 and 400 mb, the superiority of model-based forecasts over RAOBR-derived forecasts at the LOW-MID layer (850 mb level) is difficult to explain.

While results for NG,RAOBR comparisons (Table 17) suggest that icing predictors (except TESS(YALL)) perform somewhat better overall when using reduced radiosonde data compared to NOGAPS, a more definitive measure of performance difference between the two data sources is available through statistical testing. In general, tests performed on NG,RAOBR Tau = 0, 12 and 24 hr comparisons (Appendix B) do not indicate significant differences in forecasting skill (V) between model- and RAOBR-based predictions. Indeed, only one NG,RAOBR skill score difference (for TESS(YALL) at Tau = 24 hr, and favoring NG) is significant at the 95% confidence level, none at the 0.01 level of significance. These results, when combined with those drawn from Table 16 and 17 statistics, suggest that inaccuracies in model temperature and moisture data had only a limited influence upon the forecasting ability of the AIRF, NAWAU, RAP and TESS icing algorithms.

The degree of importance of vertical data resolution to

algorithm icing prediction may be examined by directly comparing NG,RAOB and NG,RAOBR statistical measure totals given in Tables 15 and 17, respectively. Model-based predictions serve as a good "fixed" standard to compare differences in RAOB- and RAOBR-based performance since NG statistics presented in both Tables 15 and 17 are largely based on identical icing predictions, with the only difference an additional 35 (or about 5% more) NOGAPS-derived predictions included with Table 15. For both forecasting capability (PF) and skill (V), 'best' statistical value totals for RAOB-derived forecasts show a significant decrease when icing predictions are based on reduced (i.e., standard level) data. The effect of radiosonde resolution on false alarm rates is quite dramatic, as evidenced by a pronounced shift in 'best' overall FAR statistics between NG,RAOB and NG,RAOBR comparisons. Specifically, while false alarm rates for NOGAPS-derived forecasts are decidedly better than those for full radiosonde (RAOB) data (FAR totals 76 to 16), icing forecasts based on reduced radiosonde data (RAOBR) have overall better false alarm rates than model-derived predictions (52 'best' values for RAOBR, 16 for NG). Collectively, results for all three statistical measures indicate that data vertical resolution plays a critical role in icing prediction performance.

For a more definitive assessment of the role of vertical data resolution in icing prediction, one may compare results of statistical tests performed on NG,RAOB and, NG,RAOBR, skill score differences. As previously discussed, many (18) significant and highly significant differences in skill score were found for NG,RAOB comparisons (all but one favoring RAOB), while only one

significant V difference (at the 95% confidence level) was obtained for NG,RAOBR comparisons. Given the largely invariant nature of the NG data for these two comparison datasets (NG,RAOB; NG,RAOBR), differences between them (as indicated by statistical significance test results) can best be attributed to differences in radiosonde vertical data resolution. This particular result strongly suggests that better (i.e., higher) vertical data resolution is a very important factor for enhanced icing prediction performance.

For two icing algorithms (AIRF and TESS), comparisons of NOGAPS and radiosonde-based forecasts in terms of icing type and intensity are possible. For each of four icing agreement classifications, Table 18 gives the cumulative number of icing predictors, summed over Tau = 0, 12 and 24 hr, with best percent frequency of agreement based on NOGAPS and radiosonde (RAOB) data, at selected icing layers. Here, the maximum possible sum (NG + RAOB) for a given layer and agreement classification is 9 (3 forecast taus x 3 icing predictors). Table 18 is based on percent frequency of agreement (NG,RAOB) comparison statistics presented as Appendix C. For all four icing classes, Table 18 'best' percentage totals for RAOB-based predictions are considerably higher than those for model-derived forecasts. The advantage of RAOB data over NOGAPS is most (least) noticeable for agreement in intensity (type). Interestingly, of the only 4 NG 'best' percentages for agreement in intensity, 3 correspond to the data-deficient HIGH layer. Concerning forecast/observed agreement in type, NOGAPS-derived forecast performance is some-

Table 18. For selected icing agreement classifications, the cumulative number of icing predictors, summed over Tau = 0, 12 and 24 hr, with best percent frequency of agreement based on NOGAPS (NG) and radiosonde (RAOB) data, for selected icing layers.

	TYPES AGREE		INTENSITIES AGREE		BOTH TYPE AND INTENSITY AGREE		NEITHER TYPE NOR INTENSITY AGREE	
DATA SOURCE	NG	RAOB	NG	RAOB	NG	RAOB	NG	RAOB
LAYER								
LOW	0	0	0	0	0	0	0	0
LOW-MID	5	2	0	8	2	7	0	8
MIDDLE	0	9	0	9	0	9	0	9
MID-HIGH	3	6	0	9	0	6	3	6
HIGH	0	5	3	1	0	4	2	2
COLUMN	6	3	1	8	3	6	3	6
TOTAL	14	25	4	35	5	32	8	31

Table 19. For selected icing agreement classifications, the cumulative number of icing layers, summed over Tau = 0, 12 and 24 hr, with best percent frequency of agreement based on NOGAPS (NG) and radiosonde (RAOB) data, for selected icing predictors.

	TYPES AGREE		INTENSITIES AGREE		BOTH TYPE AND INTENSITY AGREE		NEITHER TYPE NOR INTENSITY AGREE	
DATA SOURCE	NG	RAOB	NG	RAOB	NG	RAOB	NG	RAOB
PREDICTOR								
AIRF	0	11	0	12	2	11	0	13
TESS(YALL)	9	5	1	11	3	9	7	6
TESS(Y100)	5	9	3	12	0	12	1	12
TOTAL	14	25	4	35	5	32	8	31

what better than radiosonde-based at two levels - LOW-MID and COLUMN. The superiority of full radiosonde data (standard + significant levels) over NOGAPS is most striking at the MIDDLE (700 mb) layer, where all four agreement classifications have a RAOB to NG ratio of 9 to zero.

For the AIRF and two TESS predictors, the cumulative number of icing layers with 'best' percent frequency of agreement for NOGAPS and RAOB-based type/intensity predictions is presented in Table 19. Here, any selected (NG + RAOB) sum may total up to 18 (3 forecast taus x 6 icing layers). Of these three predictors, AIRF appears to benefit most from the use of radiosonde data for icing type and intensity prediction, with all four RAOB to NG 'best' percent agreement ratios strongly favoring RAOB-based forecasts. Comparison results for TESS(YALL) indicate that, for two of the 4 categories (types agree, neither type nor intensity agree), 'best' percent frequency of agreement sums are actually higher for model-derived forecasts. Interestingly, while NOGAPS data appears somewhat advantageous over radiosonde data to TESS(YALL) for differentiating icing type, the same is not true for icing intensity, where TESS(YALL) RAOB-based forecasts appear to be decidedly superior.

For selected type/intensity classifications, Table 20 gives the cumulative number of icing predictors (summed over three forecast lengths) with 'best' percent frequency of agreement based on NOGAPS and reduced-radiosonde (RAOBR) data, at selected icing layers. With the exception of forecast/observed agreement in 'both type and intensity', 'best' percentage totals favor NOGAPS-based predictions over those using reduced radiosonde

Table 20. For selected icing agreement classifications, the cumulative number of icing predictors, summed over Tau = 0, 12 and 24 hr, with best percent frequency of agreement based on NOGAPS (NG) and reduced radiosonde (RAOBR) data, for selected icing layers.

	TYPES AGREE		INTENSITIES AGREE		BOTH TYPE AND INTENSITY AGREE		NEITHER TYPE NOR INTENSITY AGREE	
DATA SOURCE	NG	RAOBR	NG	RAOBR	NG	RAOBR	NG	RAOBR
LAYER								
LOW	0	0	0	3	0	0	0	0
LOW-MID	7	1	3	3	2	3	6	2
MIDDLE	6	3	1	3	0	3	6	3
MID-HIGH	7	1	3	3	0	6	9	0
HIGH	2	2	7	0	1	0	5	0
COLUMN	8	1	6	2	5	3	9	0
TOTAL	30	8	20	14	8	15	35	5

Table 21. For selected icing agreement classifications, the cumulative number of icing layers, summed over Tau = 0, 12 and 24 hr, with best percent frequency of agreement based on NOGAPS (NG) and reduced radiosonde (RAOBR) data, for selected icing predictors.

	TYPES AGREE		INTENSITIES AGREE		BOTH TYPE AND INTENSITY AGREE		NEITHER TYPE NOR INTENSITY AGREE	
DATA SOURCE	NG	RAOBR	NG	RAOBR	NG	RAOBR	NG	RAOBR
PREDICTOR								
AIRF	6	6	6	10	3	9	8	5
TESS(YALL)	13	2	7	4	2	6	13	0
TESS(Y100)	11	0	7	0	3	0	14	0
TOTAL	30	8	20	14	8	15	35	5

data. The advantage of NG over RAOBR is most pronounced for agreement in type and the category ' neither type nor intensity agree'; for the latter, all 9 'best' agreement percentages at the MID-HIGH and COLUMN layers favor model-based prediction. Although agreement in intensity comparisons show a clear advantage of NG over RAOBR at the HIGH layer (7 to 0), a lack of sufficient data (see Appendix D) precludes a statistically sound conclusion to be drawn.

Table 21 presents (NG,RAOBR) comparisons of icing type/intensity forecast ability for the AIRF and two TESS predictors. For TESS(Y100), predictions based on NOGAPS data hold a very strong advantage over those derived from reduced radiosonde data. Specifically, when all four icing agreement classes are combined, the NG to RAOBR 'best' percent frequency of agreement ratio for TESS(Y100) is an overwhelming 35 to zero. The predictor TESS(YALL) also provides an overall advantage to NOGAPS-based predictions, although the category for agreement in 'both type and intensity' favors RAOBR-derived forecasts. AIRF indicates no clear superiority in icing type and intensity prediction using either model or reduced radiosonde data. For this predictor, 2 of the 4 agreement categories favor RAOBR, one favors NG, and one (for type agreement) favors neither.

The very significant impact of vertical data resolution on icing type and intensity prediction can be noted by comparing (NG,RAOB) and (NG,RAOBR) 'best' agreement totals of Tables 19 and 21, respectively. For all three predictors, a pronounced shift in overall forecast advantage toward NOGAPS occurs as the vertical

resolution of radiosonde-based predictions drops from standard plus significant levels to essentially standard levels. For the TESS(Y100) predictor, all four of the type/intensity categories undergo a dramatic reversal in forecasting advantage (from radiosonde to model) upon diminution of radiosonde resolution. Finally, a comparison of NG versus RAOBR 'best' percent frequencies of agreement for statistical measures and icing type/intensity classifications (Tables 17 and 21, respectively) suggests that, while inaccuracies in this study's model data were overall somewhat detrimental to algorithmic (except TESS(YALL)) prediction of icing occurrence, such inaccuracies actually abetted the TESS algorithm's limited forecast ability in differentiating icing type and intensity.

6. SUMMARY AND CONCLUSIONS

As a prelude to selection of an icing potential product for inclusion into an aviation support environmental display suite under development at NRL's Marine Meteorology Division, four different aircraft icing algorithms are evaluated using meteorological data from the Navy Operational Global Atmospheric Prediction System (NOGAPS) model. The algorithms are RAP (developed by NCAR's Research Applications Program), NAWAU (used operationally by the National Centers for Environmental Prediction Aviation Weather Center), AIRF (the operational routine for the Air Force Global Weather Center) and TESS, an applications program within the Navy's Tactical Environmental Support System. The NAWAU and TESS algorithms are each evaluated as two separate predictors. NAWAU(1) requires either category 1 or 2 icing for a positive icing forecast; NAWAU(2) only considers category 2 icing

forecasts as positive. The TESS(YALL) predictor considers any non-zero icing probability as a "yes" forecast; much more restrictive, TESS(Y100) requires a 100% probability for positive icing. Verification of the algorithms is accomplished by comparing model-derived analyses and short-range forecasts to pilot reports (PIREPs) of aircraft icing; for the AIRF and TESS routines, statistical comparisons include icing type and intensity. The sensitivity of icing algorithms to model data accuracy is examined by comparing model-derived predictions with coincidental forecasts based on radiosonde data. The impact of data vertical resolution on forecast performance is assessed through comparisons of icing predictions based on full radiosonde data (standard plus significant levels) with coincidental predictions based on standard level data.

For verification purposes, some 1750 pilot reports over the continental U.S. during the spring of 1995 were utilized. The largest concentration of reports (including those of moderate or greater intensity) was over the lower Great Lakes and northern Ohio Valley. Secondary maxima of reports (and positive icing events) were located over the central Rockies and Pacific Northwest. Negative icing reports dominated south of 35 ^{o} N, and are heavily concentrated over the south-central states. The vast majority of positive icing reports occurred between 1.5 and 5.5 km (5 to 18 thousand feet); at higher levels, negative icing reports dominated. Rime icing was the most prevalent type, indicated in about 3 of every 4 positive reports. Almost half of all icing occurrences were reported at a light intensity; moderate

icing (LGT-MDT and MDT categories) was indicated in about 2 of 5 icing events. The percent frequency of occurrence of rime icing was found to decrease somewhat with increasing intensity, while an opposite trend was noted for clear icing. For all three icing types (RME, CLR and MXD), percent frequencies of occurrence for any specified intensity category were quite similar. Finally, comparisons drawn among this study's PIREP dataset and various other observational datasets found in the literature indicate good agreement regarding the most commonly reported intensity and type of aircraft icing, namely, light rime.

For the most part, differences in icing forecast performance among algorithms are closely related to differences in temperature and moisture thresholds used by individual algorithms to infer icing. The TESS algorithm's unique computational processing negatively impacted its ability to correctly predict positive icing occurrences immediately above the surface or within layers located above surface superadiabatic conditions (most common with radiosondes). Collectively, Tau= 0, 12 and 24 hr statistical results for NOGAPS-based predictions indicate overall best forecast capability and skill (PF and V scores, respectively) for the NAWAU(1) and RAP predictors within the lower troposphere (below ~ 14000 ft MSL) and for the TESS(YALL) predictor at mid-tropospheric levels (14 to 27 thousand feet MSL). For this study's LOW-MID icing layer (centered at 850 mb, or ~ 5000 ft MSL), average (Tau = 0, 12 and 24 hr) prefigurance and skill score values exceed 0.7 and 0.5, respectively, for all three of these predictors. In general, icing predictors which indicate better forecast capability and skill at a particular

atmospheric level also have larger false alarm rates. AIRF was the most conservative predictor among the four icing algorithms, characterized by low false alarm rates but below average ability to forecast positive icing occurrences. Although statistical tests indicate that, for any given predictor, differences in skill scores as a function of forecast lead time are generally not significant, some improvement in both forecast capability and skill at mid-tropospheric levels (this study's MID-HIGH and HIGH layers) occurs for the TESS(YALL) and AIRF predictors over the 0 to 24 hr forecast interval. This improvement in forecast ability is likely related to a steadily increasing (over time) positive moisture bias found in NOGAPS between 700 and 400 mb.

Overall results indicate that the AIRF and TESS algorithms' ability to differentiate icing type and intensity is clearly limited. For this study's heavily-sampled MIDDLE icing layer (7 to 14 thousand ft MSL), only about 1 of 20 AIRF or TESS predictions correctly specified both icing type and intensity while, about 3 of every 4 such predictions could specify neither type nor intensity. Both icing algorithms show overall better ability to forecast icing type than intensity. In general, AIRF outperforms TESS(YALL) in specification of icing intensity, while the opposite holds for icing type. Both AIRF and TESS significantly underforecast rime icing. Light icing is greatly underforecast by the TESS algorithm, while the opposite occurs for severe icing. Statistical results indicate that a prediction of 'light rime' icing is generally more accurate than a discriminant icing type and intensity prediction available from either AIRF or TESS. In

comparisons between 'LGT RME' and TESS(YALL) predictions, differences in percent frequency of agreement of observed and forecast icing type and intensity (as measured by selected agreement classifications) often exceeded 20% and, in several instances (mostly at Tau = 24 hr), exceeded 30%. The fact that a prediction of a most likely category of icing (viz., light rime), based on historical PIREP datasets, may be superior to any arbitrary AIRF or TESS prediction is quite significant, especially when one considers the large computational effort expended by these algorithms (especially TESS) in providing a prediction of icing type and intensity.

Study results indicate that data vertical resolution plays a critical role in icing prediction performance. Comparisons between model-derived predictions (based on standard level + 925 mb data) and radiosonde predictions using full (standard plus significant level) data indicate that icing algorithms have decidedly better forecast capability and skill, but also higher false alarm rates, when based on the higher resolution radiosonde data. Inaccuracies in model (NOGAPS 3.4) temperature and moisture data are found to have only a limited impact upon the forecasting ability of the icing algorithms evaluated herein, with differences in forecast skill (V) between model and reduced radiosonde (based on only standard data) predictions generally not significant. Comparisons of model and reduced-radiosonde icing predictions indicate better overall forecast performance for the AIRF, NAWAU and RAP algorithms when using radiosonde data; however, for the TESS(YALL) predictor, such comparisons often yield higher PF and V values for model-derived forecasts.

Interestingly, TESS predictions of icing type based on NOGAPS data proved superior to comparable predictions using reduced radiosonde data.

Based on two important selection criteria - forecast skill and computational efficiency - the NAWAU icing algorithm seems the best choice, among the four icing algorithms evaluated in this study, for inclusion into NRL's aviation support product suite. NAWAU is computationally the most simple of the algorithms, using only two dozen lines of computer code to predict two probabilities (categories) of icing occurrence. Unlike the TESS and AIRF algorithms, NAWAU is not computationally burdened with unreliable attempts at inferring icing type and intensity. Overall verification statistics indicate that NAWAU ranks high in both forecast capability and skill, slightly ahead of RAP. Forecast skill for TESS is comparable to (or perhaps slightly better than) NAWAU, but only if all nonzero icing probabilities are considered definitive 'yes' forecasts (as per TESS(YALL)). In an operational setting, a TESS(YALL) forecast skill level would not be obtained, since many forecasters would likely not issue a definitive positive forecast for icing if presented guidance indicating only a 10 or 20% chance for trace rime or mixed icing.

For the operational user, a highly desirable feature of any icing potential graphical display is its ability to accurately depict areal coverage of icing. Unfortunately, temperature- and humidity-based prediction algorithms applied to rather coarse numerical model data (such as 2.5° NOGAPS) usually specify regions of potential icing well in excess of actual areal extent.

For best delineation of icing regions (and improved forecast ability), the highest possible resolution model data should be utilized for icing predictions. Within the Navy, such data could be supplied by COAMPS (Coupled Ocean/Atmosphere Mesoscale Prediction System) (Hodur, 1993), a high resolution model presently being readied for operational implementation at FNMOC. Additionally, the use of satellite data to screen cloud-free regions out of algorithm-derived icing potential threat areas would be beneficial (Thompson et al., 1997). Operationally, any combined model/satellite icing product would require additional time and computational demands over a 'model only' product and would be subject to uncertainties regarding satellite data availability.

REFERENCES

Ahnert,P.R., 1991: Precision and comparability of National Weather Service upper air measurements. Preprints, Seventh Symposium on Meteorological Observations and Instrumentation, Amer. Meteor. Soc., 227-231.

Air Weather Service, 1980: Forecaster's guide on aircraft icing. AWS/TR-80/001, U.S. Air Force, Scott AFB, IL.

Brown, B.G., G.Thompson, R.T.Bruintjes, and R.Bullock, 1994: Verification of in-flight icing algorithms: Results of the WISP94 real-time icing prediction and evaluation program (WRIPEP). Research Applications Program, National Center for Atmospheric Research, Boulder, CO.

Cohen, I.D., 1983: Analysis of AFGL aircraft icing data. AFGL-TR-83-0170, Air Force Geophysics Laboratory, Hanscom AFB, MA.

Cornell, D., C.A.Donahue, and C.Keith, 1995: A comparison of aircraft icing forecast models. AFCCC/TN-95/004, Air Force Combat Climatology Center, Scott AFB, IL.

Forbes, G.S., Y.Hu, B.G.Brown, B.C.Bernstein, and M.K.Politovich, 1993: Examination of conditions in the proximity of pilot reports of aircraft icing during STORM-FEST. Proceedings, Fifth International Conference on Aviation Weather Systems, Amer. Meteor. Soc., 282-286.

Goerss,J.S., and P.A.Phoebus, 1993: The multivariate optimum

interpolation analysis of meteorological data at the Fleet Numerical Oceanography Center. NRL/FR/7531--92-9413, Naval Research Laboratory, Monterey, CA.

Hanssen , A.W., and W.J.A. Kuipers, 1965: On the relationship between the frequency of rain and various meteorological parameters. Koninklijk Nederlands Meteorologisch Instituut, Meded. Verhand., 81, 2-15.

Hodur,R.M., 1993: Development and testing of the Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). NRL/MR/7533--93-7213, Naval Research Laboratory, Monterey, CA.

Hogan,T., T.E.Rosmond, and R.Gelaro, 1991: The NOGAPS forecast model: A technical description. NOARL Report 13, Naval Research Laboratory, Monterey, CA.

Jeck, R.K., 1985: A new characterization of the icing environment below 10000 ft AGL from 7000 miles of measurements in supercooled clouds. Proceedings, Seventh Annual Workshop on Meteorological and Environmental Inputs to Aviation Systems, NASA Conference Publication 2388.

Knapp, D.I., 1992: Verification report: Comparison of various icing analysis and forecast techniques. Air Force Global Weather Center, Offutt AFB, NE.

Naval Oceanographic Office, 1988: Tactical Environmental Support System baseline (TESS 2.0) program performance specification Volume 1. TESS 88-04 V-1, Environmental Systems Office, Stennis

Space Center, Bay St. Louis, MS.

Perkins, P.J., W.Lewis, and D.R. Mulholland, 1957: Statistical study of aircraft icing probabilities at the 700- and 500- millibar levels over ocean areas in the Northern Hemisphere. NACA Tech. Note 3984.

Schultz, P., and M.K. Politovich, 1992: Toward the improvement of aircraft icing forecasts for the continental United States. Wea. Forecasting, 7, 492-500.

Thompson, G., R.T.Bruintjes, and B.G.Brown, 1995: A comprehensive icing prediction and evaluation program. Preprints, Sixth Conference on Aviation Weather Systems, Amer. Meteor. Soc., 243-248.

Thompson, G., T.F. Lee, and R. Bullock, 1997: Using satellite data to reduce spatial extent of diagnosed icing. Accepted by Wea. Forecasting.

Thompson, J.K., 1955: 1954 Icing presentations for major commands. WADC Tech. Note WCT 55-26, Wright Air Development Center, Wright-Patterson AFB, OH.

APPENDIX A - Statistical Measures -
Model and Radiosonde Comparisons

TAU = 0 HR		LAYER					
		LOW	LOW-MID	MIDDLE	MID-HIGH	HIGH	COLUMN
NO.PIREPS(Y/N)		1/20	24/57	89/90	44/91	5/82	136/20
				PREFIGURANCE (PF)			
AIRF	(NG)	0.0	0.500	0.236	0.318	0.600	0.596
	(RAOB)	0.0	0.667	0.708	0.750	0.600	0.890
NAWAU(1)	(NG)	0.0	0.875	0.685	0.318	0.0	0.838
	(RAOB)	0.0	0.875	0.910	0.818	0.200	0.971
NAWAU(2)	(NG)	0.0	0.667	0.584	0.045	0.0	0.691
	(RAOB)	0.0	0.833	0.809	0.432	0.0	0.890
RAP	(NG)	0.0	0.875	0.697	0.250	0.0	0.853
	(RAOB)	0.0	0.875	0.910	0.705	0.0	0.971
TESS(YALL)	(NG)	0.0	0.625	0.371	0.568	0.600	0.801
	(RAOB)	1.0	0.875	0.584	0.545	0.800	0.632
TESS(Y100)	(NG)	0.0	0.583	0.213	0.295	0.400	0.471
	(RAOB)	0.0	0.792	0.506	0.500	0.400	0.603
				FALSE ALARM RATE (FAR)			
AIRF	(NG)	0.0	0.053	0.122	0.110	0.061	0.100
	(RAOB)	0.050	0.088	0.289	0.187	0.024	0.150
NAWAU(1)	(NG)	0.050	0.123	0.256	0.154	0.0	0.250
	(RAOB)	0.050	0.140	0.411	0.330	0.037	0.450
NAWAU(2)	(NG)	0.050	0.105	0.211	0.044	0.0	0.150
	(RAOB)	0.050	0.140	0.289	0.187	0.0	0.150
RAP	(NG)	0.050	0.123	0.256	0.143	0.0	0.200
	(RAOB)	0.050	0.140	0.411	0.264	0.012	0.300
TESS(YALL)	(NG)	0.0	0.105	0.200	0.253	0.098	0.250
	(RAOB)	0.050	0.263	0.367	0.297	0.049	0.250
TESS(Y100)	(NG)	0.0	0.070	0.111	0.077	0.012	0.200
	(RAOB)	0.050	0.123	0.267	0.176	0.024	0.150
				HK DISCRIMINANT (V)			
AIRF	(NG)	0.0	0.447	0.114	0.208	0.539	0.496
	(RAOB)	-0.050	0.579	0.419**	0.563**	0.576	0.740
NAWAU(1)	(NG)	-0.050	0.752	0.430	0.164	0.0	0.588
	(RAOB)	-0.050	0.735	0.499	0.489**	0.163	0.521
NAWAU(2)	(NG)	-0.050	0.561	0.373	0.001	0.0	0.541
	(RAOB)	-0.050	0.693	0.520	0.245	0.0	0.740
RAP	(NG)	-0.050	0.752	0.441	0.107	0.0	0.653
	(RAOB)	-0.050	0.735	0.499	0.441**	-0.012	0.671
TESS(YALL)	(NG)	0.0	0.520	0.171	0.315	0.502	0.551
	(RAOB)	0.950	0.612	0.218	0.249	0.751	0.382
TESS(Y100)	(NG)	0.0	0.513	0.102	0.219	0.388	0.271
	(RAOB)	-0.050	0.669	0.239	0.324	0.376	0.453

** difference at 0.01 level of significance
* difference at 0.05 level of significance

TAU = 12 HR		LAYER					
		LOW	LOW-MID	MIDDLE	MID-HIGH	HIGH	COLUMN
NO.PIREPS(Y/N)		1/22	24/63	94/97	46/100	6/87	142/20
				PREFIGURANCE (PF)			
AIRF	(NG)	0.0	0.208	0.149	0.435	0.333	0.570
	(RAOB)	0.0	0.667	0.723	0.739	0.333	0.908
NAWAU(1)	(NG)	0.0	0.708	0.691	0.413	0.167	0.873
	(RAOB)	0.0	0.875	0.926	0.783	0.0	0.979
NAWAU(2)	(NG)	0.0	0.625	0.596	0.022	0.0	0.683
	(RAOB)	0.0	0.833	0.830	0.413	0.0	0.908
RAP	(NG)	0.0	0.708	0.691	0.304	0.0	0.859
	(RAOB)	0.0	0.875	0.926	0.674	0.0	0.979
TESS(YALL)	(NG)	0.0	0.667	0.372	0.717	0.333	0.824
	(RAOB)	1.0	0.875	0.596	0.543	0.500	0.627
TESS(Y100)	(NG)	0.0	0.375	0.255	0.391	0.167	0.479
	(RAOB)	0.0	0.792	0.521	0.478	0.167	0.599
				FALSE ALARM RATE (FAR)			
AIRF	(NG)	0.0	0.048	0.103	0.100	0.046	0.150
	(RAOB)	0.045	0.079	0.299	0.170	0.023	0.150
NAWAU(1)	(NG)	0.045	0.095	0.247	0.200	0.0	0.250
	(RAOB)	0.045	0.127	0.423	0.320	0.034	0.450
NAWAU(2)	(NG)	0.045	0.048	0.175	0.040	0.0	0.100
	(RAOB)	0.045	0.127	0.309	0.170	0.0	0.150
RAP	(NG)	0.045	0.079	0.227	0.140	0.0	0.200
	(RAOB)	0.045	0.127	0.423	0.250	0.011	0.300
TESS(YALL)	(NG)	0.0	0.079	0.175	0.320	0.149	0.250
	(RAOB)	0.045	0.238	0.381	0.300	0.057	0.250
TESS(Y100)	(NG)	0.0	0.032	0.103	0.110	0.023	0.200
	(RAOB)	0.045	0.111	0.289	0.170	0.023	0.150
				HK DISCRIMINANT (V)			
AIRF	(NG)	0.0	0.161	0.046	0.335	0.287	0.420
	(RAOB)	-0.045	0.587**	0.424**	0.569*	0.310	0.758*
NAWAU(1)	(NG)	-0.045	0.613	0.444	0.213	0.167	0.623
	(RAOB)	-0.045	0.748	0.503	0.463*	-0.034	0.529
NAWAU(2)	(NG)	-0.045	0.577	0.420	-0.018	0.0	0.583
	(RAOB)	-0.045	0.706	0.521	0.243*	0.0	0.758
RAP	(NG)	-0.045	0.629	0.465	0.164	0.0	0.659
	(RAOB)	-0.045	0.748	0.503	0.424*	-0.011	0.679
TESS(YALL)	(NG)	0.0	0.587	0.197	0.397	0.184	0.574
	(RAOB)	0.955	0.637	0.214	0.243	0.443	0.377
TESS(Y100)	(NG)	0.0	0.343	0.152	0.281	0.144	0.279
	(RAOB)	-0.045	0.681*	0.233	0.308	0.144	0.449

** difference at 0.01 level of significance
* difference at 0.05 level of significance

TAU = 24 HR		LAYER					
		LOW	LOW-MID	MIDDLE	MID-HIGH	HIGH	COLUMN
NO.PIREPS(Y/N)		1/21	23/59	95/94	46/96	7/84	141/21
				PREFIGURANCE (PF)			
AIRF	(NG)	0.0	0.087	0.200	0.478	0.286	0.589
	(RAOB)	0.0	0.652	0.726	0.717	0.429	0.894
NAWAU(1)	(NG)	0.0	0.739	0.642	0.457	0.143	0.858
	(RAOB)	0.0	0.870	0.905	0.783	0.143	0.972
NAWAU(2)	(NG)	0.0	0.652	0.579	0.0	0.0	0.652
	(RAOB)	0.0	0.870	0.811	0.435	0.0	0.901
RAP	(NG)	0.0	0.739	0.632	0.283	0.0	0.837
	(RAOB)	0.0	0.870	0.905	0.674	0.0	0.972
TESS(YALL)	(NG)	0.0	0.696	0.347	0.826	0.571	0.844
	(RAOB)	1.0	0.913	0.589	0.543	0.571	0.631
TESS(Y100)	(NG)	0.0	0.478	0.221	0.370	0.143	0.504
	(RAOB)	0.0	0.826	0.526	0.500	0.286	0.624
				FALSE ALARM RATE (FAR)			
AIRF	(NG)	0.0	0.051	0.106	0.125	0.060	0.190
	(RAOB)	0.048	0.085	0.309	0.167	0.012	0.143
NAWAU(1)	(NG)	0.095	0.085	0.266	0.177	0.0	0.286
	(RAOB)	0.048	0.136	0.436	0.312	0.024	0.429
NAWAU(2)	(NG)	0.095	0.068	0.149	0.031	0.0	0.143
	(RAOB)	0.048	0.136	0.319	0.156	0.0	0.143
RAP	(NG)	0.095	0.085	0.266	0.146	0.0	0.238
	(RAOB)	0.048	0.136	0.436	0.240	0.012	0.286
TESS(YALL)	(NG)	0.0	0.068	0.170	0.292	0.119	0.286
	(RAOB)	0.048	0.237	0.383	0.281	0.048	0.238
TESS(Y100)	(NG)	0.0	0.051	0.085	0.125	0.036	0.143
	(RAOB)	0.048	0.119	0.298	0.177	0.012	0.143
				HK DISCRIMINANT (V)			
AIRF	(NG)	0.0	0.036	0.094	0.353	0.226	0.398
	(RAOB)	-0.048	0.567**	0.418**	0.551	0.417	0.751*
NAWAU(1)	(NG)	-0.095	0.654	0.376	0.279	0.143	0.572
	(RAOB)	-0.048	0.734	0.469	0.470	0.119	0.543
NAWAU(2)	(NG)	-0.095	0.584	0.430	-0.031	0.0	0.510
	(RAOB)	-0.048	0.734	0.491	0.279*	0.0	0.758
RAP	(NG)	-0.095	0.654	0.366	0.137	0.0	0.599
	(RAOB)	-0.048	0.734	0.469	0.434*	-0.012	0.686
TESS(YALL)	(NG)	0.0	0.628	0.177	0.534*	0.452	0.558
	(RAOB)	0.952	0.676	0.206	0.262	0.524	0.393
TESS(Y100)	(NG)	0.0	0.427	0.136	0.245	0.107	0.361
	(RAOB)	-0.048	0.707	0.228	0.323	0.274	0.481

** difference at 0.01 level of significance
* difference at 0.05 level of significance

APPENDIX B - Statistical Measures -
Model and Reduced Radiosonde Comparisons

TAU = 0 HR		LAYER					
		LOW	LOW-MID	MIDDLE	MID-HIGH	HIGH	COLUMN
NO.PIREPS(Y/N)		1/12	24/56	84/90	37/89	5/81	125/20
				PREFIGURANCE (PF)			
AIRF	(NG)	0.0	0.500	0.238	0.243	0.600	0.576
	(RAOBR)	1.0	0.250	0.298	0.189	0.400	0.392
NAWAU(1)	(NG)	0.0	0.875	0.679	0.270	0.0	0.824
	(RAOBR)	0.0	0.875	0.655	0.432	0.0	0.872
NAWAU(2)	(NG)	0.0	0.667	0.571	0.054	0.0	0.688
	(RAOBR)	0.0	0.833	0.536	0.135	0.0	0.704
RAP	(NG)	0.0	0.875	0.690	0.297	0.0	0.840
	(RAOBR)	0.0	0.875	0.631	0.378	0.0	0.864
TESS(YALL)	(NG)	0.0	0.625	0.393	0.541	0.600	0.808
	(RAOBR)	0.0	0.375	0.321	0.622	0.600	0.584
TESS(Y100)	(NG)	0.0	0.583	0.226	0.243	0.400	0.456
	(RAOBR)	0.0	0.250	0.298	0.459	0.200	0.488
				FALSE ALARM RATE (FAR)			
AIRF	(NG)	0.0	0.054	0.122	0.112	0.062	0.100
	(RAOBR)	0.0	0.054	0.122	0.079	0.0	0.050
NAWAU(1)	(NG)	0.0	0.125	0.256	0.157	0.0	0.250
	(RAOBR)	0.0	0.107	0.278	0.157	0.0	0.250
NAWAU(2)	(NG)	0.0	0.107	0.211	0.045	0.0	0.150
	(RAOBR)	0.0	0.107	0.156	0.034	0.0	0.100
RAP	(NG)	0.0	0.125	0.256	0.146	0.0	0.200
	(RAOBR)	0.0	0.107	0.233	0.112	0.0	0.150
TESS(YALL)	(NG)	0.0	0.107	0.200	0.258	0.099	0.250
	(RAOBR)	0.0	0.071	0.189	0.191	0.012	0.150
TESS(Y100)	(NG)	0.0	0.071	0.111	0.079	0.012	0.200
	(RAOBR)	0.0	0.018	0.111	0.067	0.0	0.050
				HK DISCRIMINANT (V)			
AIRF	(NG)	0.0	0.446	0.116	0.131	0.538	0.476
	(RAOBR)	1.0	0.196	0.175	0.111	0.400	0.342
NAWAU(1)	(NG)	0.0	0.750	0.423	0.113	0.0	0.574
	(RAOBR)	0.0	0.768	0.377	0.275	0.0	0.622
NAWAU(2)	(NG)	0.0	0.560	0.360	0.009	0.0	0.538
	(RAOBR)	0.0	0.726	0.380	0.101	0.0	0.604
RAP	(NG)	0.0	0.750	0.435	0.151	0.0	0.640
	(RAOBR)	0.0	0.768	0.398	0.266	0.0	0.714
TESS(YALL)	(NG)	0.0	0.518	0.193	0.282	0.501	0.558
	(RAOBR)	0.0	0.304	0.133	0.431	0.588	0.434
TESS(Y100)	(NG)	0.0	0.512	0.115	0.165	0.388	0.256
	(RAOBR)	0.0	0.232	0.187	0.392	0.200	0.438

TAU = 12 HR		LAYER					
		LOW	LOW-MID	MIDDLE	MID-HIGH	HIGH	COLUMN
NO.PIREPS(Y/N)		1/13	24/62	89/97	38/97	6/86	130/20
				PREFIGURANCE (PF)			
AIRF	(NG)	0.0	0.208	0.157	0.368	0.333	0.554
	(RAOBR)	1.0	0.250	0.315	0.237	0.333	0.415
NAWAU(1)	(NG)	0.0	0.708	0.685	0.368	0.167	0.877
	(RAOBR)	0.0	0.875	0.674	0.395	0.0	0.877
NAWAU(2)	(NG)	0.0	0.625	0.584	0.026	0.0	0.692
	(RAOBR)	0.0	0.833	0.562	0.105	0.0	0.715
RAP	(NG)	0.0	0.708	0.685	0.342	0.0	0.862
	(RAOBR)	0.0	0.875	0.652	0.342	0.0	0.869
TESS(YALL)	(NG)	0.0	0.667	0.393	0.711	0.333	0.823
	(RAOBR)	0.0	0.375	0.337	0.579	0.500	0.562
TESS(Y100)	(NG)	0.0	0.375	0.270	0.316	0.167	0.454
	(RAOBR)	0.0	0.250	0.315	0.447	0.167	0.477
				FALSE ALARM RATE (FAR)			
AIRF	(NG)	0.0	0.048	0.103	0.103	0.047	0.150
	(RAOBR)	0.0	0.048	0.113	0.103	0.012	0.050
NAWAU(1)	(NG)	0.0	0.097	0.247	0.196	0.0	0.250
	(RAOBR)	0.0	0.097	0.278	0.155	0.0	0.250
NAWAU(2)	(NG)	0.0	0.048	0.175	0.041	0.0	0.100
	(RAOBR)	0.0	0.097	0.155	0.031	0.0	0.100
RAP	(NG)	0.0	0.081	0.227	0.144	0.0	0.200
	(RAOBR)	0.0	0.097	0.237	0.113	0.0	0.150
TESS(YALL)	(NG)	0.0	0.081	0.175	0.320	0.151	0.250
	(RAOBR)	0.0	0.065	0.196	0.186	0.023	0.150
TESS(Y100)	(NG)	0.0	0.032	0.103	0.103	0.023	0.200
	(RAOBR)	0.0	0.016	0.113	0.062	0.0	0.050
				HK DISCRIMINANT (V)			
AIRF	(NG)	0.0	0.160	0.054	0.265	0.287	0.404
	(RAOBR)	1.0	0.202	0.201	0.134	0.322	0.365
NAWAU(1)	(NG)	0.0	0.612	0.438	0.173	0.167	0.627
	(RAOBR)	0.0	0.778	0.396	0.240	0.0	0.627
NAWAU(2)	(NG)	0.0	0.577	0.409	-0.015	0.0	0.592
	(RAOBR)	0.0	0.737	0.407	0.074	0.0	0.615
RAP	(NG)	0.0	0.628	0.459	0.198	0.0	0.662
	(RAOBR)	0.0	0.778	0.415	0.229	0.0	0.719
TESS(YALL)	(NG)	0.0	0.586	0.218	0.391	0.182	0.573
	(RAOBR)	0.0	0.310	0.141	0.393	0.477	0.412
TESS(Y100)	(NG)	0.0	0.343	0.167	0.213	0.143	0.254
	(RAOBR)	0.0	0.234	0.201	0.386	0.167	0.427

TAU = 24 HR		LAYER					
		LOW	LOW-MID	MIDDLE	MID-HIGH	HIGH	COLUMN
NO.PIREPS(Y/N)		1/12	23/58	90/94	39/93	7/83	130/21
				PREFIGURANCE (PF)			
AIRF	(NG)	0.0	0.087	0.211	0.436	0.286	0.569
	(RAOBR)	1.0	0.261	0.300	0.256	0.286	0.415
NAWAU(1)	(NG)	0.0	0.739	0.633	0.359	0.143	0.846
	(RAOBR)	0.0	0.870	0.656	0.410	0.0	0.869
NAWAU(2)	(NG)	0.0	0.652	0.567	0.0	0.0	0.646
	(RAOBR)	0.0	0.870	0.567	0.128	0.0	0.723
RAP	(NG)	0.0	0.739	0.622	0.256	0.0	0.823
	(RAOBR)	0.0	0.870	0.633	0.359	0.0	0.862
TESS(YALL)	(NG)	0.0	0.696	0.367	0.795	0.571	0.831
	(RAOBR)	0.0	0.348	0.333	0.590	0.429	0.554
TESS(Y100)	(NG)	0.0	0.478	0.233	0.308	0.143	0.485
	(RAOBR)	0.0	0.217	0.311	0.462	0.143	0.485
				FALSE ALARM RATE (FAR)			
AIRF	(NG)	0.0	0.052	0.106	0.129	0.060	0.190
	(RAOBR)	0.0	0.034	0.117	0.108	0.012	0.048
NAWAU(1)	(NG)	0.0	0.086	0.266	0.161	0.0	0.286
	(RAOBR)	0.0	0.103	0.287	0.151	0.0	0.238
NAWAU(2)	(NG)	0.0	0.069	0.149	0.032	0.0	0.143
	(RAOBR)	0.0	0.103	0.160	0.032	0.0	0.095
RAP	(NG)	0.0	0.086	0.266	0.140	0.0	0.238
	(RAOBR)	0.0	0.103	0.245	0.108	0.0	0.143
TESS(YALL)	(NG)	0.0	0.069	0.170	0.280	0.120	0.286
	(RAOBR)	0.0	0.069	0.202	0.172	0.024	0.143
TESS(Y100)	(NG)	0.0	0.052	0.085	0.129	0.036	0.143
	(RAOBR)	0.0	0.017	0.117	0.065	0.0	0.048
				HK DISCRIMINANT (V)			
AIRF	(NG)	0.0	0.035	0.105	0.307	0.225	0.379
	(RAOBR)	1.0	0.226	0.183	0.149	0.274	0.368
NAWAU(1)	(NG)	0.0	0.653	0.367	0.198	0.143	0.560
	(RAOBR)	0.0	0.766	0.368	0.260	0.0	0.631
NAWAU(2)	(NG)	0.0	0.583	0.418	-0.032	0.0	0.503
	(RAOBR)	0.0	0.766	0.407	0.096	0.0	0.628
RAP	(NG)	0.0	0.653	0.356	0.117	0.0	0.585
	(RAOBR)	0.0	0.766	0.389	0.251	0.0	0.719
TESS(YALL)	(NG)	0.0	0.627*	0.196	0.515	0.451	0.545
	(RAOBR)	0.0	0.279	0.131	0.418	0.404	0.411
TESS(Y100)	(NG)	0.0	0.427	0.148	0.179	0.107	0.342
	(RAOBR)	0.0	0.200	0.194	0.397	0.143	0.437

* difference at 0.05 level of significance

APPENDIX C - Type/Intensity Categories - Model and Radiosonde Comparisons

TAU = 0 HR	NO. DATA	AIRF (NG)	AIRF (RAOB)	PREDICTOR TESS(YALL) (NG)	PREDICTOR TESS(YALL) (RAOB)	TESS(Y100) (NG)	TESS(Y100) (RAOB)
				TYPES AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	21(0,2)	14.3%	14.3%	33.3%	21.1%	33.3%	19.0%
MIDDLE	83(1,3)	15.7%	42.2%	17.1%	22.5%	9.6%	21.7%
MID-HIGH	40	22.5%	50.0%	25.0%	12.5%	2.5%	10.0%
HIGH	5	60.0%	60.0%	20.0%	60.0%	0.0%	20.0%
COLUMN	125(3,2)	35.2%	45.6%	49.2%	18.7%	16.8%	12.8%
				INTENSITIES AGREE			
LAYER							
LOW	1(0,1)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	24(0,2)	29.2%	29.2%	4.2%	31.8%	4.2%	29.2%
MIDDLE	89(1,4)	6.7%	34.8%	1.1%	8.2%	1.1%	6.7%
MID-HIGH	44	18.2%	47.7%	2.3%	13.6%	0.0%	13.6%
HIGH	5	40.0%	40.0%	20.0%	20.0%	20.0%	0.0%
COLUMN	136(3,2)	28.7%	47.1%	12.8%	12.7%	2.9%	12.5%
				BOTH TYPE AND INTENSITY AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	21(0,2)	14.3%	4.8%	0.0%	5.3%	0.0%	4.8%
MIDDLE	82(1,3)	3.7%	24.4%	0.0%	3.8%	0.0%	3.7%
MID-HIGH	40	10.0%	37.5%	2.5%	2.5%	0.0%	2.5%
HIGH	5	40.0%	40.0%	0.0%	20.0%	0.0%	0.0%
COLUMN	124(3,2)	16.1%	29.0%	9.1%	2.5%	0.8%	1.6%
				NEITHER TYPE NOR INTENSITY AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	21(0,2)	66.7%	61.9%	61.9%	47.4%	61.9%	52.4%
MIDDLE	82(1,3)	80.5%	45.1%	81.5%	72.2%	89.0%	74.4%
MID-HIGH	40	70.0%	40.0%	75.0%	77.5%	97.5%	80.0%
HIGH	5	40.0%	40.0%	60.0%	40.0%	80.0%	80.0%
COLUMN	124(3,2)	50.8%	34.7%	46.3%	70.5%	80.6%	75.8%

Note: "NO. DATA" counts pertain to AIRF and TESS(Y100) predictors; the number of data for the TESS(YALL) predictor based on NG (RAOB) data is found by subtracting the first (second) number in parenthesis from the given data count.

TAU = 12 HR	NO. DATA	AIRF (NG)	AIRF (RAOB)	PREDICTOR TESS(YALL) (NG)	PREDICTOR TESS(YALL) (RAOB)	PREDICTOR TESS(Y100) (NG)	PREDICTOR TESS(Y100) (RAOB)
				TYPES AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	21(2,2)	14.3%	14.3%	36.8%	21.1%	14.3%	19.0%
MIDDLE	87(3,3)	9.2%	43.7%	15.5%	22.6%	9.2%	21.8%
MID-HIGH	42	26.2%	45.2%	23.8%	14.3%	2.4%	9.5%
HIGH	5	40.0%	40.0%	20.0%	40.0%	0.0%	0.0%
COLUMN	129(2,2)	34.9%	45.7%	48.8%	17.3%	13.2%	12.4%
				INTENSITIES AGREE			
LAYER							
LOW	1(0,1)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	24(2,2)	8.3%	29.2%	4.5%	31.8%	0.0%	29.2%
MIDDLE	94(3,4)	7.4%	35.1%	1.1%	10.0%	1.1%	8.5%
MID-HIGH	46	17.4%	43.5%	2.2%	10.9%	0.0%	10.9%
HIGH	6	16.7%	16.7%	16.7%	16.7%	16.7%	0.0%
COLUMN	142(2,2)	23.9%	46.5%	10.7%	12.9%	2.1%	12.7%
				BOTH TYPE AND INTENSITY AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	21(2,2)	9.5%	4.8%	0.0%	5.3%	0.0%	4.8%
MIDDLE	86(3,3)	5.8%	24.4%	0.0%	3.6%	0.0%	3.5%
MID-HIGH	42	7.1%	33.3%	2.4%	2.4%	0.0%	2.4%
HIGH	5	20.0%	20.0%	0.0%	20.0%	0.0%	0.0%
COLUMN	128(2,2)	14.8%	28.9%	7.1%	2.4%	0.8%	1.6%
				NEITHER TYPE NOR INTENSITY AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	21(2,2)	85.7%	61.9%	57.9%	47.4%	85.7%	52.4%
MIDDLE	86(3,3)	88.4%	43.0%	83.1%	71.1%	89.5%	73.3%
MID-HIGH	42	64.3%	45.2%	76.2%	78.6%	97.6%	83.3%
HIGH	5	60.0%	60.0%	60.0%	60.0%	80.0%	100.%
COLUMN	128(2,2)	54.7%	34.4%	46.8%	71.4%	85.9%	75.8%

TAU = 24 HR	NO. DATA	AIRF (NG)	AIRF (RAOB)	PREDICTOR TESS(YALL) (NG)	PREDICTOR TESS(YALL) (RAOB)	TESS(Y100) (NG)	TESS(Y100) (RAOB)
				TYPES AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	20(2,2)	0.0%	10.0%	50.0%	22.2%	30.0%	20.0%
MIDDLE	88(2,4)	14.8%	43.2%	18.6%	21.4%	10.2%	20.5%
MID-HIGH	42	28.6%	42.9%	31.0%	11.9%	2.4%	9.5%
HIGH	6	33.3%	50.0%	50.0%	50.0%	0.0%	16.7%
COLUMN	128(1,2)	32.8%	43.0%	46.5%	15.9%	14.1%	12.5%
				INTENSITIES AGREE			
LAYER							
LOW	1(0,1)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	23(2,2)	8.7%	30.4%	4.8%	33.3%	4.3%	30.4%
MIDDLE	95(2,5)	9.5%	34.7%	1.1%	8.9%	1.1%	8.4%
MID-HIGH	46	21.7%	43.5%	2.2%	10.9%	0.0%	10.9%
HIGH	7	14.3%	28.6%	14.3%	14.3%	14.3%	0.0%
COLUMN	141(1,2)	24.1%	46.1%	10.7%	12.9%	2.8%	12.8%
				BOTH TYPE AND INTENSITY AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	20(2,2)	0.0%	5.0%	0.0%	5.6%	0.0%	5.0%
MIDDLE	87(2,4)	6.9%	24.1%	0.0%	3.6%	0.0%	3.4%
MID-HIGH	42	9.5%	31.0%	2.4%	2.4%	0.0%	2.4%
HIGH	6	16.7%	33.3%	0.0%	16.7%	0.0%	0.0%
COLUMN	127(1,2)	11.8%	27.6%	6.3%	2.4%	0.8%	1.6%
				NEITHER TYPE NOR INTENSITY AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	20(2,2)	90.0%	65.0%	44.4%	44.4%	65.0%	50.0%
MIDDLE	87(2,4)	81.6%	43.7%	80.0%	73.5%	88.5%	74.7%
MID-HIGH	42	61.9%	45.2%	69.0%	81.0%	97.6%	83.3%
HIGH	6	66.7%	50.0%	33.3%	50.0%	83.3%	83.3%
COLUMN	127(1,2)	53.5%	36.2%	48.4%	72.8%	84.3%	75.6%

APPENDIX D - Type/Intensity Categories -
Model and Reduced Radiosonde Comparisons

TAU = 0 HR	NO. DATA	AIRF (NG)	AIRF (RAOBR)	PREDICTOR TESS(YALL) (NG)	PREDICTOR TESS(YALL) (RAOBR)	TESS(Y100) (NG)	TESS(Y100) (RAOBR)
				TYPES AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	21(0,2)	14.3%	9.5%	33.3%	21.1%	33.3%	14.3%
MIDDLE	79(1,1)	16.5%	25.3%	17.9%	10.3%	10.1%	8.9%
MID-HIGH	34	14.7%	17.6%	26.5%	14.7%	2.9%	0.0%
HIGH	5	60.0%	40.0%	20.0%	40.0%	0.0%	0.0%
COLUMN	116(3,1)	32.8%	29.3%	46.9%	23.5%	18.1%	10.3%
				INTENSITIES AGREE			
LAYER							
LOW	1	0.0%	100.%	0.0%	0.0%	0.0%	0.0%
LOW-MID	24(0,2)	29.2%	16.7%	4.2%	4.5%	4.2%	0.0%
MIDDLE	84(1,1)	7.1%	16.7%	1.2%	1.2%	1.2%	1.2%
MID-HIGH	37	16.2%	13.5%	0.0%	8.1%	0.0%	0.0%
HIGH	5	40.0%	20.0%	20.0%	0.0%	20.0%	0.0%
COLUMN	125(3,1)	28.0%	24.0%	10.7%	8.9%	3.2%	2.4%
				BOTH TYPE AND INTENSITY AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	21(0,2)	14.3%	0.0%	0.0%	5.3%	0.0%	0.0%
MIDDLE	78(1,1)	3.8%	16.7%	0.0%	0.0%	0.0%	0.0%
MID-HIGH	34	8.8%	11.8%	0.0%	8.8%	0.0%	0.0%
HIGH	5	40.0%	20.0%	0.0%	0.0%	0.0%	0.0%
COLUMN	115(3,1)	15.7%	18.3%	7.1%	6.1%	0.9%	0.0%
				NEITHER TYPE NOR INTENSITY AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	21(0,2)	66.7%	76.2%	61.9%	78.9%	61.9%	85.7%
MIDDLE	78(1,1)	79.5%	73.1%	80.5%	88.3%	88.5%	89.7%
MID-HIGH	34	76.5%	82.4%	73.5%	85.3%	97.1%	100.%
HIGH	5	40.0%	60.0%	60.0%	60.0%	80.0%	100.%
COLUMN	115(3,1)	53.0%	65.2%	48.2%	73.7%	79.1%	87.0%

Note: "NO. DATA" counts pertain to AIRF and TESS(Y100) predictors; the number of data for the TESS(YALL) predictor based on NG (RAOBR) data is found by subtracting the first (second) number in parenthesis from the given data count.

TAU = 12 HR	NO. DATA	AIRF (NG)	AIRF (RAOBR)	PREDICTOR TESS(YALL) (NG)	PREDICTOR TESS(YALL) (RAOBR)	TESS(Y100) (NG)	TESS(Y100) (RAOBR)
				TYPES AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	21(2,2)	14.3%	9.5%	36.8%	21.1%	14.3%	14.3%
MIDDLE	83(3,1)	9.6%	26.5%	16.3%	9.8%	9.6%	8.4%
MID-HIGH	35	17.1%	17.1%	28.6%	14.3%	2.9%	0.0%
HIGH	5	40.0%	40.0%	20.0%	40.0%	0.0%	0.0%
COLUMN	119(2,1)	31.9%	30.3%	47.0%	22.0%	14.3%	10.1%
				INTENSITIES AGREE			
LAYER							
LOW	1	0.0%	100.%	0.0%	0.0%	0.0%	0.0%
LOW-MID	24(2,2)	8.3%	16.7%	4.5%	4.5%	0.0%	0.0%
MIDDLE	89(3,1)	7.9%	16.9%	1.2%	1.1%	1.1%	1.1%
MID-HIGH	38	15.8%	13.2%	2.6%	7.9%	0.0%	0.0%
HIGH	6	16.7%	16.7%	16.7%	0.0%	16.7%	0.0%
COLUMN	130(2,1)	23.1%	23.8%	9.4%	9.3%	2.3%	2.3%
				BOTH TYPE AND INTENSITY AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	21(2,2)	9.5%	0.0%	0.0%	5.3%	0.0%	0.0%
MIDDLE	82(3,1)	6.1%	17.1%	0.0%	0.0%	0.0%	0.0%
MID-HIGH	35	5.7%	11.4%	2.9%	8.6%	0.0%	0.0%
HIGH	5	20.0%	20.0%	0.0%	0.0%	0.0%	0.0%
COLUMN	118(2,1)	14.4%	18.6%	6.0%	6.0%	0.8%	0.0%
				NEITHER TYPE NOR INTENSITY AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	21(2,2)	85.7%	76.2%	57.9%	78.9%	85.7%	85.7%
MIDDLE	82(3,1)	87.8%	72.0%	82.3%	88.9%	89.0%	90.2%
MID-HIGH	35	71.4%	82.9%	71.4%	85.7%	97.1%	100.%
HIGH	5	60.0%	60.0%	60.0%	60.0%	80.0%	100.%
COLUMN	118(2,1)	57.6%	64.4%	48.3%	74.4%	84.7%	87.3%

TAU = 24 HR	NO. DATA	AIRF (NG)	AIRF (RAOBR)	PREDICTOR TESS(YALL) (NG)	PREDICTOR TESS(YALL) (RAOBR)	TESS(Y100) (NG)	TESS(Y100) (RAOBR)
				TYPES AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	20(2,2)	0.0%	10.0%	50.0%	16.7%	30.0%	10.0%
MIDDLE	84(2,1)	15.5%	26.2%	19.5%	9.6%	10.7%	8.3%
MID-HIGH	36	22.2%	16.7%	33.3%	11.1%	2.8%	0.0%
HIGH	6	33.3%	33.3%	50.0%	33.3%	0.0%	0.0%
COLUMN	119(1,1)	29.4%	30.3%	43.2%	20.3%	15.1%	10.1%
				INTENSITIES AGREE			
LAYER							
LOW	1	0.0%	100.%	0.0%	0.0%	0.0%	0.0%
LOW-MID	23(2,2)	8.7%	17.4%	4.8%	4.8%	4.3%	0.0%
MIDDLE	90(2,1)	10.0%	16.7%	1.1%	1.1%	1.1%	1.1%
MID-HIGH	39	20.5%	12.8%	2.6%	5.1%	0.0%	0.0%
HIGH	7	14.3%	14.3%	14.3%	0.0%	14.3%	0.0%
COLUMN	130(1,1)	23.1%	24.6%	9.3%	7.0%	3.1%	2.3%
				BOTH TYPE AND INTENSITY AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	20(2,2)	0.0%	0.0%	0.0%	5.6%	0.0%	0.0%
MIDDLE	83(2,1)	7.2%	16.9%	0.0%	0.0%	0.0%	0.0%
MID-HIGH	36	8.3%	11.1%	2.8%	5.6%	0.0%	0.0%
HIGH	6	16.7%	16.7%	0.0%	0.0%	0.0%	0.0%
COLUMN	118(1,1)	10.2%	19.5%	5.1%	4.3%	0.8%	0.0%
				NEITHER TYPE NOR INTENSITY AGREE			
LAYER							
LOW	0	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
LOW-MID	20(2,2)	90.0%	75.0%	44.4%	83.8%	65.0%	90.0%
MIDDLE	83(2,1)	80.7%	72.3%	79.0%	89.0%	88.0%	90.4%
MID-HIGH	36	66.7%	83.3%	66.7%	88.9%	97.2%	100.%
HIGH	6	66.7%	66.7%	33.3%	66.7%	83.3%	100.%
COLUMN	118(1,1)	55.9%	64.4%	51.3%	76.9%	83.1%	87.3%